Student thesis series INES nr 281

Comparison of MODIS-Algorithms for Estimating Gross Primary Production from Satellite Data in semi-arid Africa

Martin Nilsson

2013
Department of
Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Martin Nilsson (2013). Comparison of MODIS-Algorithms for Estimating Gross Primary Production from Satellite Data in semi-arid Africa.
Bachelor degree thesis, 15 credits in *Physical Geography and Ecosystem Science*
Department of Physical Geography and Ecosystem Science, Lund University

ABSTRACT

The climatic patterns of the world are changing and with them the spatial distribution of global terrestrial carbon; the food and fiber of the world and in itself an important factor in the changing climate. Knowledge of how the terrestrial carbon stock is changing, its distribution and quantity, is important in understanding how the patterns of the world are changing and large scale models using remotely sensed data have emerged for this purpose. This study compares four vegetation related MODIS (Moderate Resolution Imaging Spectroradiometer) products, derived from MODIS satellite data using algorithms which calculates the two vegetation indices, Normalized Difference Vegetation Index (*NDVI*) and Enhanced Vegetation Index (*EVI*), and the two biophysical factors, Leaf Area Index (*LAI*) and absorbed Fraction of Photosynthetically Active Radiation (*FPAR*). The comparison is in their ability to estimate intra-annual variations of Gross Primary Production (GPP); this is done using the time-series data of quality screened eddy covariance (EC) Flux Tower stations from the Carbo Africa network as truth data.

The results show a modest agreement between the different vegetation metrics and EC Flux Tower derived GPP, with an overall average coefficient of determination (R^2) of 0.63 for LAI, R^2 of 0.51 for NDVI, R^2 of 0.52 FPAR and a R^2 of 0.49 for EVI, using all stations and years of data. When each station received the same weight, i.e. using the correlation of all observation for each station and then calculating the average, the overall correlation improved, still showing LAI as the best predictor of Flux Tower GPP with a R^2 of 0.62, but with an improved EVI with a R^2 of 0.61, while NDVI and FPAR had an R^2 of 0.57 and 0.59 respectively. This result and the observed large variation in between stations, e.g. NDVI between an R^2 of 0.62 and 0.83 for the station Demokeya compared to an R^2 of 0.32 and 0.49 of NDVI for the station Tchizalamou, may indicate a site specific proficiency of the vegetation metrics. When the observations within the growing period were tested separately a strong decrease in correlation was observed, with an average R^2 between 0.41 – 0.56 for all station and years and an average R^2 between 0.36 – 0.45 for all sites using all observations for each station regardless of year, lending strength to the assumption that the non-vegetation period observations affect the correlation greatly.

The study concludes that up scaling of an intra-annual standardized major axis regression model based solely on the relationship between any of these metrics and Flux Tower estimated GPP is inadvisable due to the modest overall intra-annual agreement between the metrics and GPP. It is also concluded that since the vegetation metrics display site specific proficiency, models of GPP would benefit from site specific ancillary data that describes vegetation-limiting factors, e.g. water availability.

ABSTRAKT

Klimatmönstren världen över förändras och med dessa den globala distributionen och mängden av landbundet kol, dvs vegetationen som bland annat nyttjas som mat och fiber. Också I sig självt en viktig faktor i klimatets utveckling genom dess roll i energi- och vattenkretsloppen. Vetskap om kvantitet och distribution av landbundet kol och hur detta förändras är en viktigt del av arbetet i att förstå hur de globala mönster förändras, och för denna avsikt har bredskaliga modeller som nyttjar satellit data framtagits. Denna studie jämför fyra vegetations relaterade MODIS (Moderate Resolution Imaging Spectroradiometer) produkter, som erhålls från MODIS satellit data genom algoritmer som kalkylerar de två vegetation indexen, Normalized Vegetation Index (*NDVI*) och Enhanced Vegetation Index (*EVI*), och de två biofysiska faktorerna, Leaf Area Index (*LAI*) och absorbed Fraction of Photosynthetically Active Radiation (*FPAR*). Deras förmåga att uppskatta variationen av den totala primär produktionen (Gross Primary Production, GPP) över året jämförs, genom tidsserier av eddy kovarians (EC) data från flux torn ur Carbo Africa nätverket, vars tidsserie-utveckling används som sanningspunkter varmot variationen från de motsvarande tidsserierna av algoritmerna jämförs.

Resultatet visar en blygsam korrelation mellan de olika vegetations algoritmernas reslutat och EC flux torn uppskattat GPP, med ett medel av determinationskoefficienter (R^2) på 0.49 för EVI, 0.51 för NDVI,0.52 för FPAR och ett R^2 på 0.63 för LAI, då data från alla stationer och år användes. När var station erhöll lika stor vikt, dvs då korrelationen kalkylerades för samtliga observationer från var station, varpå medel togs fram, förbättrades korrelationen över lag. fLAI visades fortfarande som den bästa prediktorn av flux-torns uppskattad GPP med ett R^2 på 0.62, ett starkt förbättrat R^2 för EVI på 0.61erhölls, medans NDVI och FPAR visa ett R^2 på 0.57 respektive 0.59. Detta resultat och en stundtals stor variation mellan stationer, t.ex. NDVI med ett R^2 mellan 0.62 och 0.83 för stationen Demokeya jämfört med ett R^2 mellan 0.32 och 0.49 för NDVI och stationen Tchizalamou, visar kanske på plats specifika förmågor hos vegetations algoritmerna. När observationer inom vegetationsperioden testades separat observerades en starkt minskad korrelation, med ett medel R^2 mellan 0.41 – 0.56 för alla stationer och år, och ett R^2 mellan 0.36 – 0.45 för alla platser vid användning av samtliga observationer för varje station oberoende av år, vilket indikerar att observationerna utanför växtperioden har stort inflytande på korrelationen.

En slutsats av studien är att uppskalning av en standrardiserad storaxels regressions modell för inom annuell variation baserad endast på relationen mellan en av dessa vegetations algoritmer och flux-torns uppskattad GPP ej är att rekommendera med tanke på den blygsamma överrensstämmelsen mellan vegetationsalgoritmerna och flux torn uppskattat GPP. En annan slutsats är att eftersom dessa vegetations algoritmer uppvisar plats specifika förmågor skulle modeller av GPP ha fördel av plats specifik stöd-data som beskriver faktorer som begränsar vegetation, t.ex. vattentillgänglighet.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to extend my outmost gratitude to the principal investigators for the sites Agoufou, Bontioli, Demokeya, Malopeni, Maun, Mongu, Skukuza and Tchizalamou, described further in section 1.4 of this study, for providing data and approving the use of said data. Without you this study would not have been possible!

While the opportunity is presented I would also like to thank my supervisor Thomas Holst and the research director at my institution Jonas Ardö, for insights and support.

Thank you all!

1. INTRODUCTION

According to the Intergovernmental Panel on Climate Change (Solomon *et al.* 2007) the global temperatures are on the rise showing an increasing linear trend during the last 100 years (1906-2005) of 0.74 °C and this rise in temperature is predicted to influence patterns of precipitation. This trend is believed to largely result from increased concentrations of greenhouse gases of which CO_2 is considered to be the most influential, and has increased in concentration mostly through fossil fuel burning and land use change (Solomon *et al.* 2007). The change in temperature and precipitation patterns is likely to change patterns of terrestrial productivity, which may have a large impact on the availability of resources like timber and crops (Sjöström *et al.* 2013). Parry *et al.* (2007) highlighted Africa as particularly vulnerable to the impacts of climate change and estimated that between 75 to 250 million Africans would be exposed to increased water stress as a result of climate change by the year 2020 and that by the same year an estimated decrease in the yield of rain fed agriculture by up to 50% is likely to occur, which is identified in the report as a substantial part of the agricultural production. By 2080 the semi-arid and arid areas in Africa are likely to increase by 5-8 percent according to several of the different posed climate scenarios (Parry *et al.* 2007).

Aside from being a source of sustenance vegetation also plays an important climatic role, both with the direct effect of vegetation in that it changes albedo and moisture in the microclimate but also with the indirect effect of reduced CO_2 in the atmosphere due to carbon fixation (Chapin *et al.* 2011). Therefore changes in vegetation cover and quantity thereof could lend important clues as to the future resource availability and climate patterns.

The eddy covariance (*EC*) flux measuring technique has become a standard for measuring e.g. CO_2 fluxes between land and atmosphere, which can be translated into an estimate of the assimilation of carbon by the vegetation, but it suffers from a discrete spatial extent combined with a small footprint (Baldocchi *et al.* 2001). Since the African continent contains a low frequency of stations providing meteorological data (Brown 2008), which can also be said for the stations measuring CO_2 using eddy covariance (here forth called Flux Towers, presently at most 18 functioning stations across Africa), remotely sensed techniques for collecting vegetation data has been of great importance and the remotely sensed data can be used in models of large scale estimation of plant primary production (Sjöström 2012). The EC technique, while subject to spatial limitations, can be used to validate models that are using remotely sensed data by acting as ground truth points.

This study uses the framework of Hashimoto *et al.* (2012) where Flux Tower derived Gross Primary Production (*GPP*) data, a measure of carbon assimilation on the ecosystem scale, is used as ground truth points for exploring MODIS (*Moderate Resolution Imaging Spectroradiometer*) derived products, in this case the MOD09A1 and MOD15A2 sets. From the MOD09A1 set the vegetation indices (*VI's*) Normalized Difference Vegetation Index (*NDVI*, Rouse *et al.* 1973) and Enhanced Vegetation Index (*EVI*, Huete *et al.* 1997) are calculated, and from the MOD15A2 set the parameters Leaf Area Index (*LAI*) and Fraction of Photosynthetically Active Radiation that is absorbed (*FPAR*, sometimes referred to as *fAPAR*) are extracted; these will hereafter be collectively referred to as vegetation metrics and their abbreviations will in the context of this study, section 1.2 and 1.3 excluded, refer to the respective MODIS products. The VI's NDVI and EVI are used to embellish the spectral characteristics of vegetation out of satellite reflectance data, LAI describes plant canopies and FPAR describes the portion of incoming light energy which the plants utilize; a more thorough description of the vegetation metrics can be found in section 1.2. While Hashimoto *et al.* (2012) evaluate these vegetation metrics for different forest sites across the globe this study does so for a series of semi-arid sites in Africa. Also this study does not delve into an annual analysis as Hashimoto *et al.* (2012) due to time constraints and sample size. For an evaluation of the MODIS17A2 set using these sites and time periods I refer to the work of Sjöström *et al.* (2013).

Following this introduction to the study, section 1.1 states the objectives of the study. Section 1.2 provides a short theoretical background and a review of past research relevant for this study follows in

section 1.3. Section 1.4 describes the study area. Section 2 describes the data used in the analysis, how the data was acquired and the specifications of the data, along with a short discussion of uncertainties with the used data sets. Following the data description the methodology used is covered in section 3. The results of the analysis will be presented in section 4 and discussed in section 5. Section 6 concludes this study with the outcome of the discussion.

1.1 Objectives

The aim of this study is to compare MODIS algorithms, i.e. the vegetation related MODIS products NDVI, EVI, LAI and FPAR, in their ability to trace the intra-annual variability of GPP across several African sites (*section 1.4, study area*), by comparing time-series of these algorithms against time-series of ground measurements (*reference points*) derived from eddy covariance Flux Tower measurements; part of the Carbon Africa project, *see section 2.1*. The goal with the comparison is to justify the construction of simple regression models estimating intra-annual GPP using the relationship between the vegetation metrics and the Flux Tower GPP.
The study also has two sub-objectives:

1. To observe the variability in between sites, evaluate whether the proficiency of these vegetation metrics are site dependent and how well they function over a wide category like savanna, i.e. if it is possible to generalize a derived regression model to such a scale.
2. To compare the vegetation metrics in their ability to handle the strong seasonality that is characteristic of these semi-arid ecosystems.

1.2 Theoretical background

Of interest to this study is how well NDVI, EVI, LAI and FPAR traces the intra-annual variations of Gross Primary Production (*GPP*), the sum of the net photosynthesis by all photosynthetic tissue at the ecosystem scale, as estimated by EC Flux Towers. Therefore this section focuses on the concept GPP and its governing factors, the basics of the EC method, remote sensing, the MODIS sensor and the above mentioned vegetation metrics, the biophysical connection between these vegetation metrics and GPP, and functional relationships in between the vegetation metrics are discussed.

Photosynthesis, the constituent of GPP, is on the cellular level dependent on the availability of photosynthetic reactants, which in turn is governed by site biota, parent material, climate and time. This since these factors govern available soil resources and plant functional types as well as and by extension the direct controls of GPP, e.g. leaf area, nitrogen content, season length, temperature, light, CO_2 and water; the direct controls act as limiting factors of GPP, as per the hypothesis of co-limitation, sometimes referred to as the functional convergence hypothesis (Chapin *et al.* 2011; Hashimoto *et al.* 2012). Hence the spatial and temporal variations in GPP adhere to the site specific biophysical properties that partake in photosynthesis, of which e.g. Leaf area index (*LAI*) and absorbed fraction of photosynthetically active radiation (*FPAR*) are measures of (Chapin *et al.* 2011). LAI describes plant canopies via a ratio, which is the one-sided leaf area per unit ground area, often expressed as $m^2\ m^{-2}$. The biophysical relationship with GPP can thereby be explained by the observation that larger canopies possess more photosynthetically active tissue, i.e. allows for more carbon to be assimilated. FPAR measures the proportion of photosynthetically active radiation (*PAR*) that is absorbed by the canopy. The biophysical relationship to GPP is then observed in the role of radiation energy in the process of photosynthesis. There are some differences in the Light Use Efficiency (*LUE*) in between different vegetation types, adding another dimension to the spatial variation of GPP (Chapin *et al.* 2011).

Eddy Covariance measures and calculates vertical turbulent fluxes of e.g. CO_2 or water vapor within an atmospheric boundary layer along with the vertical wind fluctuations using quick gas analyzers in the case of CO_2 and water vapor and using sonic anemometers for the vertical wind. These high frequency measuring tools must take at least 10 measurements per second (10 Hz). The EC measurements in this study uses 30 minute periods after which the average of all wind and CO_2 measurements are calculated,

these averages of wind and CO_2 are then used to calculate time series of the fluctuations from those averages of each measurement from the 30 minute period. The mean product between each corresponding wind and CO_2 fluctuation is then multiplied with the density of the air. This gives a notion to whether CO_2 is predominantly taken up by the ecosystem or released (Verma 1990). These measurements are taken around the clock and is an estimate of Net Ecosystem Exchange (*NEE*) which by convention is defined as CO_2 flux from the ecosystem to the atmosphere (i.e. uptake of CO_2 by the ecosystem gives a negative NEE, Chapin *et al.* 2011). Using the nighttime NEE data the R_{eco} (ecosystem respiration), which is the sum of the autotrophic and heterotrophic respiration of the ecosystem, is estimated using a temperature-dependent model whereby the relationship $GPP = NEE + R_{eco}$ gives the Gross Primary Production (Reichstein *et al.* 2005).

Remote sensing refers to the acquisition of information from afar, most commonly using aerial or satellite based sensors. When satellite based, the sensors use the atmospheric windows, the spectrum of wavelength able to pass through the atmosphere without too high extinction, to measure the reflected and emitted radiation that reaches it from the surface of the earth or from the ocean. Most often data is collected using several wavelength bands, a band referring to a sensor designed to measure how much of a certain part of the spectrum an object reflects or emits (Campbell 2006). After the acquisition of the satellite imagery follows the act of interpreting the information, when the digital numbers (*DN*) values, or rather from DN derived radiance or reflectance, of the satellite image are related to real world objects and characteristics thereof. Each object reflects solar radiation across the spectrum in a unique way which is expressed as the spectral curve or spectral signature of the object, reflected or emitted radiation as a function of wavelength, creating in a sense a spectral footprint. Vegetation has a very distinct such spectral curve, reflecting more in some regions, e.g. the Near Infrared (*NIR*, 0.7 - 1.2 µm) part of the electromagnetic spectrum. These curves differ with vegetation type and also the condition thereof, and thereby e.g. vegetation stress can be viewed. Different vegetation indices, e.g. NDVI and EVI, seek to exploit and enhance these characteristics to make vegetation more distinct in the satellite image (Campbell 2006). More specifically they utilize the contrast between the NIR radiation response to mesophyll tissue, which is great reflectance, and the energy absorption by the vegetation in the red part of the spectra (~0.6-0.7 µm) by chlorophyll to distinguish the vegetation part of the signal measured by the sensor. Hence the outcome of using a VI is governed by the photosynthetically active biomass at the ground, suggesting a relationship to GPP, albeit affected to a varying degree by the light's interaction with other ground objects as well as with the atmosphere (Campbell 2006). In the case of EVI the blue band (~0.4-0.5 µm) is also used for its recognized abilities to give a notion of atmospheric interference (Myneni *et al.* 1997). In conclusion VIs are calculated directly from reflectance data without any bias or assumptions regarding ground conditions, allowing for intra- and inter-annual comparison of phenological and biophysical parameters (Huete *et al.* 2002).

The Normalized Difference Vegetation Index (*NDVI*, Rouse *et al.* 1973) is one of the most widely used of the vegetation indices and uses the near infrared band in combination with the red band:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \qquad (1)$$

NDVI, while proven effective in many cases, seem to saturate when canopy is too dense and when canopy is to sparse the influence of the soil background is an issue (Sjöström 2012). However NDVI has a long history due to the AVHRR (*Advanced Very High Resolution Radiometer*) sensor and as such provides valuable data for long-term analysis (Huete *et al.* 2002).

The Enhanced Vegetation Index (*EVI*) on the other hand tries to account for atmospheric influence (e.g. scattering and refraction) and variable soil background reflectance, i.e. the varying reflection originating from the soil:

$$EVI = G \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + (C_1 \times \rho_{red} - C_2 \times \rho_{blue}) + L} \qquad (2)$$

Where G = 2.5 is the gain factor, C_1 = 6 and C_2 = 7.5 are band-specific atmospheric resistance correction coefficients, and L = 1 is a background correction factor (Huete *et al.* 1997). The spatial and temporal variations of VIs arise from a series of vegetation properties, e.g. leaf area, canopy structure, plant type, land cover type etcetera, and the effects of background and other non-vegetative features resulting in unwanted artifacts (Huete *et al.* 2002). Notably linked to the governing factors of GPP mentioned before. There are many other vegetation indices, like the simple ration (*SR*) and notably indices that use the Near Infrared (*NIR*) and Short Wave Infrared (*SWIR*) bands, e.g. *LSWI – Land Surface Water Index* (Xiao *et al.* 2003), using the SWIR bands ability to capture water, recognized as an essential control factor for photosynthesis (Chapin *et al.* 2011).

The Moderate Resolution Imaging Spectroradiometer (*MODIS*) is carried on the Terra satellite, launched December 18th 1999, and has provided continuous data since the end of February year 2000. It has a 16 day repeat cycle but due to its swath width it is able to provide a global data set every 1-2 days. It has a pass over time between 10:00-11:00. MODIS collects data via 36 spectral channels among which 7 are of special interest for vegetation and land surfaces; blue band (ρ_{blue} = *459-479nm*), green band (ρ_{green} = *545-565nm*), red band (ρ_{red} = *620-670nm*), near infrared 1 band (ρ_{nir} = *841-875nm*), near infrared 2 band (ρ_{nir2} = *1230-1250nm*), shortwave infrared 1 band (ρ_{swir1} = *1628-1652nm*), shortwave infrared 2 band (ρ_{swir2} = *2105-2155nm*). The amount of bands of the MODIS sensor and its onboard calibration system has made it usable for a wider range of applications and makes it more accurate than the AVHRR sensor whose NDVI time-series thus far has been one of the most used for mapping vegetation change (Campbell 2006). With the launch of the Terra satellite came a whole series of MODIS derived products, for example the sets used in this study, the MOD09A1 (raw reflectance) and MOD15A2 (*LAI* and *FPAR*) sets (Internet source: MOD09; MOD15a).

Using the raw reflectance MODIS product (MOD09A1) the NDVI and EVI is calculated as per equation 1 and 2 above. The two other MODIS products (from MOD15A2) being compared in this study, LAI and FPAR, are as previously mentioned biophysical variables involved with the canopy structure and functional processes of plants. However it should be noted that the LAI and FPAR used in this study is derived from satellite data via algorithms using look up tables. While the biophysical connection to GPP for LAI and FPAR is quite well understood the process of extracting or estimating LAI and FPAR out of reflectance data is still a complex and uncertain science which holds potential for improvement (Fensholt *et al.* 2004).

Intuitively, the above written suggests that there should be some connection between VIs and these biophysical properties and research has shown that there is a strong functional relationship between NDVI and FPAR for several vegetation types which is linear (Asrar *et al.* 1984; Myneni *et al*. 1997; Fensholt *et al.* 2004) and a functional relationship between NDVI and LAI that is exponential, or rather one of sums of the exponential functions describing radiative transfer where the photon count decays exponentially during its interaction with a canopy of leaves (Myneni *et al*. 1997; Hashimoto *et al*. 2012). However the functional relationship NDVI/FPAR is not always entirely linear and Fensholtet *et al.* (2004) point out some external factors, e.g. atmospheric influence and view angle geometry, as well as some canopy related factors, e.g. leaf angle distribution (LAD), canopy heterogeneity, soil-canopy reflectance interactions and senescent material in the canopy, that affect this linear relation. The exponential relationship between NDVI and LAI is usually described by Beer's law, through which also FPAR and LAI are correlated, since the canopy of leaves govern the fraction of absorbed PAR (Hashimoto *et al.* 2012). The relationship between LAI and NDVI has been shown to vary considerably between cover types (Myneni *et al.* 1997). These relationships are often used and VI's are often used as intermediaries in the assessment of biophysical parameters such as just LAI and FPAR (Asrar *et al.* 1984; Huete *et al.* 2002), for example the MOD15A2 set uses them when the main algorithm fails (Myneni *et al.* 1997), more on the main algorithm in the data description section 2.2.

1.3 Past Research – a review

The use of remote sensing for mapping of vegetation increased during the years following the launch of the first Landsat system satellite (*ERTS*, later Landsat MSS, 1972). A milestone in the field of satellite remote sensing of vegetation was the project to map the vegetation on the Great Plains of the USA, where researchers encountered the influence of the solar zenith angle on the ability to map vegetation, a find from which stems NDVI (Rouse *et al.* 1973). By the mid-eighties a series of vegetation indices (*VI's*) had been proposed and an increasing amount of studies in the topic performed, notably Tucker (1979) who investigated infrared properties of vegetation and Perry and Lautenschlager (1984) comparing different VI's and concluding that there were few practical differences between the VI's implemented at the time. During the eighties the AVHRR sensor was increasingly used in the remote study of vegetation coverage, and with its large coverage and NDVI data it enabled researchers to study phenological variations over large regions; among the first to do so was Reed *et al.* (1994) who collected four years of cloud free AVHRR data for which they calculated NDVI for each pixel. During research with the increasingly large AVHRR set problems were identified, e.g. the varying influence of soil background reflectance, and as a response new VI's emerged. Like the Soil Adjusted Vegetation Index (*SAVI*, Huete 1988) and EVI (Huete *et al.* 1997), which both emerged due to the noted influence of soil brightness variations in areas with sparse canopies, which is especially influential in semi-arid regions (Sjöström 2012).

With the MODIS sensors in orbit the possibilities of remote sensing vegetation is greater than ever and in more recent research a branch stemming from the noted global warming trend has emerged (Solomon *et al.* 2007) which is large scale modeling of GPP to view changes in the global carbon stock (Xiao *et al.* 2003). In the light of this many models of GPP has emerged, either multi-parameter models (e.g. VPM, TG; Xiao *et al.* 2004; Sims *et al.* 2008) or simpler models utilizing the relationship between satellite derived VI's or biophysical parameters, e.g. the vegetation metrics evaluated in this study, and GPP (Hashimoto *et al.* 2012; Sjöström *et al.* 2011). With the emergence of this increasing amount of models, the need for methods to validate them has increased; of which eddy covariance Flux Tower measurements has been used in many studies, e.g. Hashimoto *et al.* (2012), Sjöström *et al.* (2009; 2011), Xiao *et al.* 2003, Sims *et al.* 2008, Turner *et al.* 2006.

Hashimoto *et al.* (2012) evaluated the short and long term abilities of NDVI, EVI, LAI, and FPAR to estimate Flux Tower measured GPP for forest ecosystems (*deciduous, evergreen and tropical forest*). They found that among these vegetation metrics, overall, EVI most highly correlated to Flux Tower GPP with a coefficient of determination (R^2) of 0.55 for 16-day composite and 0.54 for 32-day composite period (Hashimoto *et al.* 2012, *table 3*). They also showed that the MODIS GPP product, that uses the satellite derived parameter FPAR with meteorological data (*e.g. Vapor Pressure Deficit, VPD*) to estimate GPP, explains seasonal variations in Flux Tower GPP ($R^2 = 0.68$ *for 8-day composites*) better than the above vegetation metrics; which they argue supports the need for input of meteorological data for more accurate capturing of seasonal photosynthetic variations. This need for meteorological data in models of photosynthesis is supported by other research, e.g. Running *et al.* (1989). Hashimoto *et al.'s* (2012) results also show that NDVI and FPAR saturates at values of 0.75 and above. While they refer to the correlation between these vegetation metrics and Flux Tower GPP as being modest, they conclude that EVI is useful for short term analysis of tower-estimated GPP variations, but they note that caution is warranted when using the EVI/GPP relationship since it seemingly is not constant across forest types. They also conclude that while LAI explains the intra-annual variations modestly, due to its inability to respond to short term stresses, it is the best indicator of annual GPP (R^2 *between 0.78 – 0.88*) (Hashimoto *et al.* 2012).

Privette *et al.* (2002) studied the performance of the first year MODIS LAI product and concludes that it does well for semi-arid woodland and savannas. Fensholt *et al.* (2004) evaluated the MODIS LAI and FPAR products for three semi-arid sites and showed that the products captured the seasonal dynamics well and only slightly overestimated these parameters, LAI by 2-15% and FPAR by 8-20%.

Sjöström *et al.* (2009) showed that for a semi-arid area in Africa (*Demokeya, view table 1 in this study, section 1.3*) TIMESAT-adjusted, TIMESAT is a software package that amongst other features filters or fits smooth functions to time-series of satellite data (Jönsson and Eklundh, 2002, 2004), MODIS EVI and

NDVI displayed good correlation with Flux Tower GPP (*R^2 of 0.93 and 0.90 respectively*). They point out that NDVI is sensitive to differences in soil background reflectance resulting from varying canopy coverage and saturates at high biomass. They also discuss the effects of varying solar zenith angles on vegetation indices (shown to be considerable for LAI between 0.25 and 2) and that this could affect the seasonal proficiency of the vegetation indices to predict GPP. Sjöström *et al.* (2009) conclude that applicability of vegetation indices for estimation of intra-annual variation can be greatly reduced due to these effects, especially for semi-arid regions where vegetation is sparse, which furthers the discussion of the site dependence of EVI (Sims *et al.* 2006). They also showed a slight improvement in performance when taking station footprint, i.e. the area measured by the Flux Towers, into consideration.

For semi-arid landscapes Sjöström *et al.* (2011), expanding on the study by Sjöström *et al.* (2009) of EVI to a larger set of African sites, i.e. Demokeya, Maun, Mongu, Skukuza, Tchizalamou, Wankama Fallow and Wankama Millet, showed that TIMESAT-adjusting MODIS EVI improves its ability to estimate intra-annual variations in GPP, from $R^2 = 0.61$ to $R^2 = 0.67$ for all sites and cases, and from $R^2 = 0.49$ to $R^2 = 0.57$ for all sites using vegetation period data only (*vegetation period determined when GPP reached a certain value*). Their results displayed large variation of R^2 between sites, e.g. $R^2 = 0.49$ for Skukuza and $R^2 = 0.90$ for Tchizalamou. They also found that division into narrower biome groups, namely grass dominated (*Demokeya, Tchizalamou*) and tree dominated savannas (*Maun, Mongu*), improved the ability of EVI to follow seasonal changes in Flux Tower GPP; $R^2 = 0.86$ for grass dominated and $R^2 = 0.87$ for tree dominated. They show that inclusion of evaporative fraction (EF, *equation 3*), which is considered water sufficiency, and PAR with EVI improved the correlation with Flux Tower measured GPP at the majority of the sites used in the study.

$$EF = \frac{LE}{LE + H} \tag{3}$$

Where LE is the latent heat and H is the sensible heat.

Which they argue shows the importance of including a factor controlling water availability in GPP models for dry ecosystems. Although they name the correlation between $EVI \times EF \times PAR$ and Flux Tower GPP modest, they concluded that EVI is a good predictor of intra-annual Flux Tower GPP whose performance increased when EF and PAR was introduced (Sjöström *et al.* 2011).

In a synthesis of Hickler *et al.* (2005), Sjöström *et al.* (2009; 2011; 2013), Sjöström (*2012*) concludes that EVI follows the seasonal dynamics of Flux Tower GPP at the site scale. But also that the EVI/GPP relationship varies greatly between the African sites and it may be difficult to scale up models of GPP based solely on EVI as factors that limit growth must be taken into account.

Sjöström (2012) also notes, in the relation to primary production models, that constant recalibration is needed as the models are time and place specific and do not take into account global and regional variations in e.g. vegetation type, solar radiation, soil water and temperature.

Asrar *et al.* (1984) found that, for a given biome, FPAR was linearly related to NDVI and curvi-lineraly related to LAI.

In an evaluation by Huete *et al.* (2002) of the radiometric and biophysical performance of EVI as calculated from the *Moderate Resolution Imaging Spectroradiometer* (*MODIS*) sensor indicated that EVI is sensitive to canopy variations.

1.4 Study Area

The study area is confined to a series of Flux Tower sites in Africa and their biomes; these are described shortly in Table 1 and visualized in Figure 1.

Table 1: Given are: site names and abbreviations, locations (Lat/Long, decimal degrees), general ecosystems as per the IGBP ecosystem classification (DBF: deciduous broadleaf forest, GRA: grassland, SAV: savanna), MAP - mean annual precipitation, MAT - mean annual temperature, years with data, number of weeks with data and period during which the majority of the rainfall occurs. As collected from http://gaia.agraria.unitus.it/home/sites-list (Accessed on the 13th of December 2012).

Name	Lat	Lon	Ecosystem (IGBP)	MAP (mm)	MAT (ºC)	Years with data	Weeks with data	Period with majority of rainfall
Agoufou (ML-AGG)	15.343	-1.481	GRA	374	30.2	2007	34	June - September
Bontioli (BF-BON)	10.866	-3.073	SAV	926	26.1	2004, 2006	43	May - September
Demokeya (SD-DEM)	13.283	30.478	SAV	320	26.0	2007-2009	109	June - October
Malopeni (ZA-MAP)	-23.833	31.214	SAV	458	22.2	2009	41	November - March
Maun (BW-MA1)	-19.914	23.560	SAV	464	22.0	2000-2001	83	December - March
Mongu (ZM-MON)	-15.435	23.253	DBF	945	24-26	2007-2009	90	November - March
Skukuza (ZA-KRU)	-25.020	31.497	SAV	547	21.9	2000-2008	339	November - March
Tchizalamou (CG-TCH)	-4.289	11.656	SAV[1]	1150	25.7	2006-2007	70	October - April

[1] For details on the IGBP ecosystem classification, see Lambin and Geist (2006).

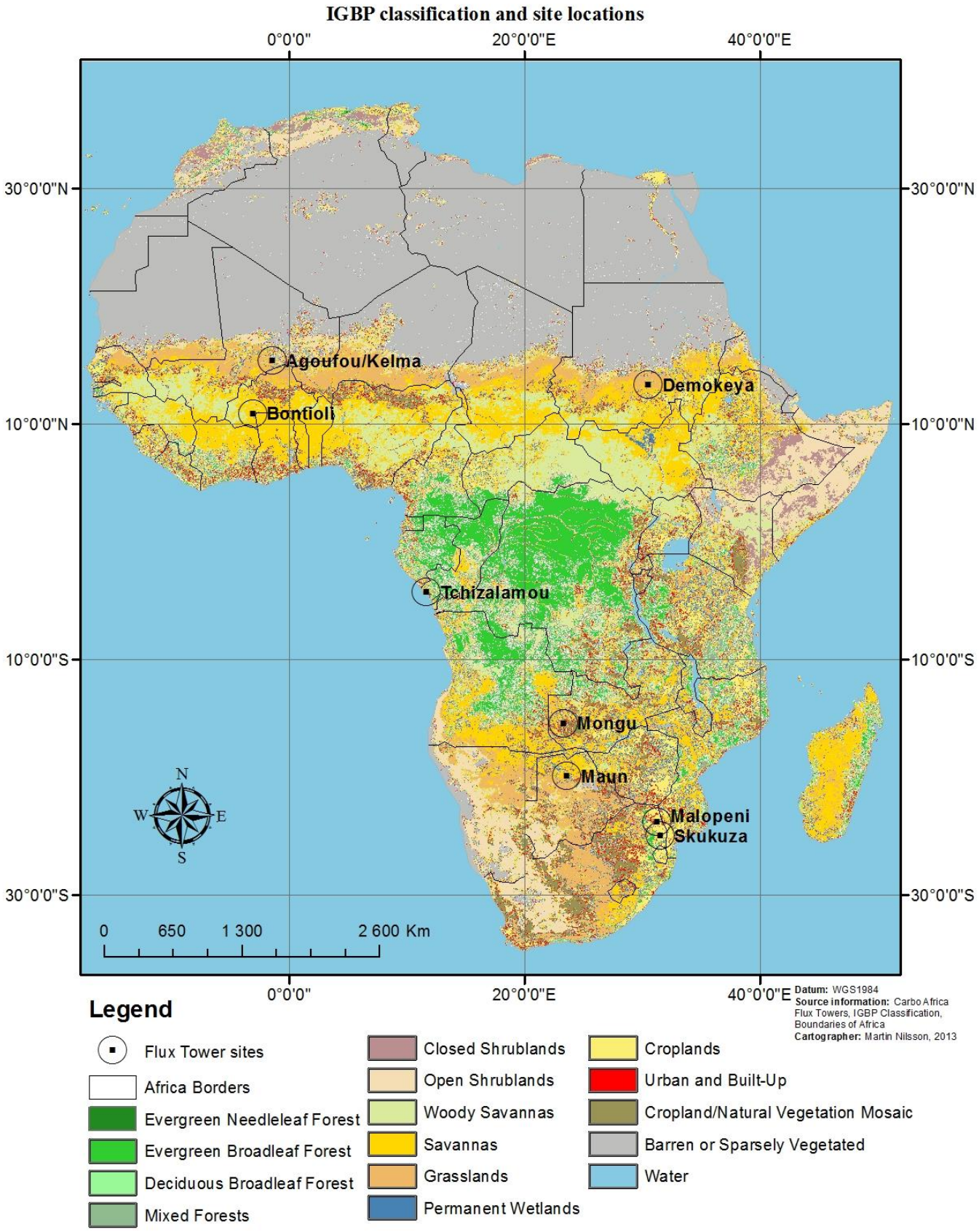

Figure 1*:* Map showing the Flux tower sites used in the study, along with an IGBP land cover classification.

2. DATA DESCRIPTION

Two methods of data acquisition were used for this study, the first was the eddy covariance technique used to measure fluxes and estimate GPP (*data from Flux Towers*) at point locations, and the second was a remote sensing method resulting in satellite derived MODIS products, aimed at identifying vegetation greenness or biophysical parameters that concern photosynthesis and monitor changes thereof.
In this section the data sets are described, how they have been acquired and how they are derived, what their strengths and weaknesses are, and what constraints they've been subjected to.

2.1 Flux Tower data - Source and information

The eddy covariance technique has become a standard for measuring fluxes of CO_2 between the land and atmosphere at the ecosystem scale (Sjöström *et al.* 2011) and with the emergence of several commercial eddy-flux systems these formerly exclusive apparatuses have become more widely used. In 2006 the Carbo Africa project was launched in an effort to coordinate and standardize EC Flux Tower sites in Africa to increase knowledge of the role Africa has in the global carbon cycle (Bombelli *et al.* 2009). But as previously stated the number of stations in the network is limited, there are only 18 sites spread over the African continent of which some are not active today, and they are discrete in space. To therefore estimate primary production over larger areas in Africa a different approach must be used, e.g. models using remotely sensed data, these can then be used in conjunction with eddy covariance data in order to calibrate the models (Sjöström *et al.* 2011).

The Flux Tower data used in this study [Table 1] was downloaded from the Carbo Afraica database (gaia.agraria.unitus.it/database/carboafrica). The data used in the analysis was strongly governed by the Flux Tower data that was readily available and the years for which this data had sufficient coverage over the vegetation period. The standardized Level 4 Carbo Africa Flux Tower data set was used which among other meteorological and environmental variables contains gap-filled NEE of CO_2 data, and from that, calculated GPP, as explained in the theoretical background. This aggregated into 8-day composites. Standardized means that the data format and the methods used on the data set follow the convention set by the Carbo Africa project. Gap-filled means that missing values in the Flux Tower time-series are filled, if possible, using the Marginal Distribution Sampling method (*MDS*, Reichstein *et al.* 2005) and the Artificial Neural Network method (*ANN*, Papale and Valentine 2003); to clarify these gap-filling methods are both used on the data separately and the Level 4 Carbo Africa set contains separate time-series for each gap-fill method. The MDS gap-filled data was chosen for this study in accordance with Hashimoto *et al.* (2012). MDS takes the average NEE value under similar meteorological conditions within a window of the missing data to replace it (Sjöström *et al.* 2013). This study used the standard storage term, i.e. were the concentration of CO_2 accumulation during non-turbulence in the layer from the ground up to the EC measuring IRGA (*Infra-red gas analyzer*) are considered constant and measured at a single point at the top of the EC tower. Whereas Hashimoto *et al.* (2012) used the original storage term, i.e. a storage term provided by the principal investigator which is either derived as the standard storage term or using a profile system where CO_2- concentrations are measured at some levels along the height of the EC tower, with the argument that it is more reliable at tall tower sites (Papale *et al.* 2006). The reason this study deviates from Hashimoto *et al.* (2012) when it comes to the used storage term is the lower availability of data using the original storage term. Since the Flux towers at the sites in this study are likely to be lower than the Flux towers for the forest sites used in Hashimoto *et al*. (2012), due to significantly lower vegetation height, the impact of this decision was deemed low.

All the Flux Tower stations where viewed in Google Earth to check areal biome homogeneity and while visual land cover classification was not performed, the content of Table 1 was checked for plausibility. Areas around stations were considered to be homogeneous for at least 1 km for all stations but for the Bontioli, Mongu and Tchizalamo sites where there were some notations.

For Bontioli the satellite imagery was taken 2009-03-09 and 1.5 km south west of the site there was an open barren area which seems to be a part of a dried up riverbed that is breaking the homogeneity of the landscape, this was however considered to be outside of the Flux Tower footprint. On the satellite imagery for the Mongu site (taken 2005-09-19) there was decidedly more vegetation on the south, south eastern side of the station (0.31 km), which notably is where the prevailing winds come from during the wet season. Tchizalamo was homogenous for 0.9 km, but notably close to the ocean (approximately 12 km) which would make it subject to a maritime microclimate. It should be noted that Sjöström *et al.* (2011) highlights the potentially heterogenic footprint of the Skukuza Flux Tower station, however the visual interpretation conducted in Google Earth for this study concluded it to be homogenous.

Flux Tower measurements are subject to many potential errors, e.g. instrumental problems and physical issues like harsh weather and different terrain phenomena (Massman and Lee 2002), but many of these can be mitigated and avoided through careful consideration when placing a station. An uncertainty with the eddy covariance method that could be an issue is that the shape and size of the Flux Tower footprint varies between sites and within sites over the year, since it depends on the surface roughness, wind direction and speed and on the height of the tower at the site (Schmid 2002), it is however assumed that the Flux Tower footprint is approximately comparable to the MODIS pixels in this study due to the homogeneity restriction imposed on the sites. Relevant to this study Archibald *et al.* (2009) describes the NEE, from which the used GPP is calculated, in semi-arid ecosystems as pulsing dependent on rainfall events and denotes that the standard gap-filling procedures are not made to account for this. Hence the gap-filled values may not be representative.

However, Flux Towers provide direct measurements over continuous temporal scales as well as ancillary data for those time periods which can be used in a series of applications. Thus, even though they are vulnerable to several issues as mentioned above, they provide a good source for validation of satellite algorithms.

2.2 MODIS data - Source and information

The MODIS data was gathered from a server hosted by the Oak Ridge National Laboratory which provides MODIS data in the text file format ASCII (*American Standard Code for Information Interchange*) precompiled for different Flux Tower field sites or allows for selection of subsets of MODIS data (ORNL DAAC 2009), courtesy of NASA. Two sets of MODIS data were acquired, the MOD09A1 (*raw reflection data*) and MOD15A2 (*LAI* and *FPAR*) sets. The MOD09A1 set was used to calculate time-series for the VI's (*NDVI* and *EVI*) and from the MOD15A2 set LAI and FPAR time-series were extracted.

MOD09A1 provides surface reflectance in 8-day composites at a 500 m resolution. In this Level 3 data, which is raw data (considered Level 0) that has been radiometrically calibrated (considered Level 1) and atmospherically corrected to yield a surface reflectance product (considered Level 2), each pixel contains the optimal L2G observation during an 8-day period, L2G is Level 2 data that has been gridded as a means of separating geolocation from compositing. In choosing the optimal L2G observation, coverage, absence of clouds or shadow thereof, aerosol loading and low view angle is considered. From this (*7 band*) raw reflectance data set the NDVI and EVI was calculated using equation (1) and (2) in contrast to Hashimoto *et al.* (2012) who used the MOD13Q1 set to extract NDVI and EVI time-series. It is presumed that the use of the NDVI and EVI equations will yield the same results as the MOD13Q1 set and it is not further investigated. The surface reflectance band quality description accompanying the MOD09A1 set (*the surf_refl_qc_500m field, 32-bit*) was not used, however the cloud state bits of the state flags (*the surf_refl_state_500m field, 16-bit*) were. For more information about the MOD09A1 set, and in particular its quality description and state flags refer to the "*MODIS Surface Reflectance User's Guide*" (Vermote *et al.* 2011, Internet Source: MOD09).

MOD15A2 provides Leaf Area Index (*LAI*, scaled to integer by a factor of 0.1) and Fraction of Photosynthetically Active Radiation (*FPAR*, scaled to integer by a factor of 0.01) composited every 8-days with a 1 km resolution (Internet source: MOD15a). The algorithm that derives the LAI and FPAR

compares the observed canopy reflectance, which is atmospherically corrected using the Bi-directional Reflectance Distribution Function (*BRDF*, Knyazikhin *et al.* 1998), against modeled radiance for a suite of canopy structures and soil patterns using biome specific lookup tables (*LUT*); and an solution is accepted if the difference is less than or equal to the corresponding uncertainty (Knyazikhin *et al.* 1998; Wang *et. al.* 2001). If there are multiple solutions the algorithm uses a weighted mean where the weights are based on the frequency of occurrence of a given solution. If the main algorithm fails, due to clouds or atmospheric effects (Wang *et. al.* 2001), a back-up algorithm sets in that uses empirical MODIS specific NDVI and LAI/FPAR relationships to produce a representation. For further information refer to the "*Algorithm Theoretical Basis Document*" for the MOD15 set (Knyazikhin *et al.* 1999, Internet Source: MOD15b). The MODIS collection 5 was used which at this date (2013-01-30) is the second newest collection, superseded by collection 6, however collection 6 is still being evaluated. A *collection* is a MODIS data archive that has been reprocessed in order to incorporate e.g. better calibration and algorithm refinements.

The main algorithm was used for LAI/FPAR in 99% of the data. As with the MOD09A1 set the MOD15A2 derived time-series were checked for cloud flags.

Satellite data comes with a series of issues, for example issues of spatial and temporal resolution, atmospheric influence and accuracy of the sensor when it comes to geolocation, i.e. the ability of the sensor to locate the real location of an object. Some of these will be touched upon below.

With decrease in spatial resolution of satellite data the pixels are likely to contain an increasing amount of radiative contribution, i.e. radiation measured by the sensor, from the background (Tian *et al.* 2000), arguably even higher in sparsely vegetated areas like in this study.

As for the geolocation accuracy of the MODIS products there is a 70% probability that the perceived object is within 50 m of the actual object (Hashimoto *et al.* 2012). Tan *et al.* (2006) reported that the average overlap between MODIS grid cells and actual observations were less than 30% due to gridding artifacts and effects of viewing geometry. Therefore direct comparison of field measurements with MODIS data becomes problematic since reflectance retrievals are not necessarily centered on the precise location of the pixel used (Sjöström *et al.* 2011). But as Sjöström *et al.* (*2011*) notes, if Flux Towers are located in a relatively large homogenous landscape, tower pixels can provide a reasonably good representation of the conditions at the vicinity of the sites.

The MODIS data quality indicator mentioned above is not taken into account which is a potential source of error. Clouds are accounted for though and the fact that the optimal L2G observation is chosen for each 8-day period reduces the risk of low quality observations.

3. METHODOLOGY

The methodology of this paper was aimed to follow that of Hashimoto *et al.* (2012) in their approach of comparing vegetation metrics to GPP gained from quality screened Flux Tower sites, thereby assessing the accuracy of the vegetation metrics abilities to estimate GPP. This is done for different ecosystems (*see section 1.4*) than Hashimoto *et al.* (2012) to extend the comparison of the vegetation metrics to a larger set of ecosystems. However there are some differences from the work of Hashimoto et al. beyond study area, most importantly the annual analysis was excluded due to time constraints and small sample size, and the ability of the MODIS 17 GPP product in estimating the intra-annual variations of Flux Tower estimated GPP has been evaluated by Sjöström *et al.* (2013) using the same tower sites and years of data that was used in this study.

3.1 Data preprocessing

The Flux Tower data, being the foremost limiter of observations, was viewed for inconsistencies and relevant temporal coverage. Time-series for the standardized level 4 MDS gap-filled GPP were then extracted. In accordance with Hashimoto *et al.* (2012) each site was visually assessed to avoid the scaling issues a heterogenic site could infer (*section 2.1*). In addition to checking site homogeneity the Flux Tower data was evaluated for temporally overlapping coverage with the MODIS data and a gap filled ratio of less than 20% was demanded for each year. Data not meeting these requirements were consequently removed. Negative GPP were set to missing data.

Time-series were extracted from the MOD09A1 set, which was then used to produce time series for NDVI and EVI, and from the MOD15A2 set LAI and FPAR time-series were directly extracted. If the MODIS data displayed less than 80% coverage over a year that year was removed from the analysis, to maximize observation pairs over the years, and if an 8-day composite was cloud flagged it was replaced using a MVC (*Maximum Value Composite*) window if possible, else set to missing data. The MVC replaced a value given that the one of the two adjacent composites contained a value; adjacent composites that were cloud flagged were not used.

3.2 Analysis

The correlation was explored using the determination coefficient (R^2) and the significance thereof was used to assess whether a linear relation between the Flux Tower time series and the different MODIS data derived time series was present, following Hashimoto *et al.* (2012). Reduced Major Axis Regression (*RMA*), or often referred to as Standardized Major Axis Regression (*SMA*), which this study will refer to it as, was used to further investigate the relationship. SMA is a regression model that unlike linear regression minimizes the sum of the product of the deviations in both x and y from the regression line (Sokal & Rohlf 1995). This regression model is often used when both variables are subject to measurement errors, i.e. it translates errors not only along the y-axis but along the x-axis as well (Sokal & Rohlf 1995). *Root mean square error* (*RMSE*) was calculated using the SMA, as one would use a linear regression model. The null hypothesis that the SMA slope was zero was also tested.

The first part of the analysis sought to find how well the different vegetation metrics explain the intra-annual variations of the Flux Tower GPP estimates, this was done calculating the R^2 between each vegetation metric and the Flux Tower GPP for each year, site and composite size (*section 4.1*), whereon the average R^2 of all years and sites depending on the vegetation metrics and composite sizes were derived. RMSE was also averaged in this manner using the results from each year and site. Since R^2 was averaged for each year and site the sample size for each year was of importance, and due to the seasonality of the sites a notion of the data covering some part of the growth season was prudent. Hence years with a sample size of less than fifteen 8-day composites and maximum Flux Tower GPP lower than 2 gC m^{-2} day^{-}

[1] were removed, this threshold was decided on from viewing the Flux Tower derived GPP of the growing season for the different sites. It was reasoned that if no coverage existed for at least some part of the vegetation period the correlation would be controlled entirely by observations not guaranteed to be affected by any sort of vegetation reflectance, observations likely subject to more noise. Setting a sample size threshold increases the likelihood that the sample correlation is representative for the true correlation. However, with a sample size of 15 in a population of 46 the margin of error is roughly 21%.

The second part of the analysis assumed that the conditions affecting the Flux Tower GPP estimates and the remotely sensed data remained somewhat constant in between the years for each site. Following that assumption the correlation for each site was sought using all the observations regardless of the time of the observation (*section 4.2*).

Under the presumption that the different savanna sites are comparable all observations for all savanna sites were pooled, whereby correlation and RMSE was calculated (*section 4.3*).

The seasonality of the sites led to the question of how well the vegetation metrics correlate to the Flux Tower GPP during the growing season, therefore the same analysis as above was conducted for observations were Flux Tower GPP indicated plant activity (GPP estimate > 0.6 gC m^{-2} day^{-1}).

Comparisons were performed for three composite sizes, 8-day (*original data*), 16-day and 32-day composites; the last two derived from the 8-day composites. The Flux Tower data 16-day composite was created by averaging the value of the two 8-day composites representing that time period of the year, and for the MODIS data the 16-day composites were created using the Maximum Value Composite method (*MVC*) on the corresponding composites representing the same time period as the 16-day composite. This implied that a Flux Tower 16-day composite was created only if there were measurements for both corresponding 8-day composites; otherwise it was marked as missing value. For the MODIS data it sufficed with one of the 8-day composite having coverage. The same principle was applied in the creation of the 32-day composites, but for the Flux Tower data only three 8-day composites with coverage was needed for a composite to be created. For the MODIS data two 8-day composites with values were considered sufficient. The last 32-day composite of the year, even though they're treated later in the analysis as a full 32-day period, covers only two 8-day composites and were consequently derived with the same premises as a 16-day composites would.

To handle outlier values in the MODIS data the mean value and the standard deviation for each year of data was calculated, each value that deviated more than three standard deviations from the mean were then removed. This was not done for the Flux Tower data since it for several of the sets would remove the vegetation peak and adjacent composite periods.

4. RESULTS

The seasonality of the sites and that all vegetation metrics followed the trend of the Flux Tower measured GPP was clearly observed in the data [Figure 2], however what was also apparent was the variation in correlation between the Flux Tower measured GPP and the different vegetation metrics for each site and year [Table 2]. For example for the Skukuza-site a high variability in performance between years was observed, e.g. a minimum coefficient of determination (R^2) of 0.07 and a maximum R^2 of 0.79 for EVI and a minimum R^2 of 0.28 and a maximum R^2 of 0.89 for LAI, for the 8-day composites, this variation generally persisted when observations outside of vegetation period were removed. An example of the variability is shown in Figure 3 where EVI is compared between the Demokeya site for year 2008 and the Tchizalamou site for year 2007 showing the difference between sites and the LAI is compared in between year 2001 and year 2002 for the site Skukuza. For Mongu and Tchizalamou a low overall performance was observed, with an average maximum R^2 of 0.45 and 0.44, respectively, for the different vegetation metrics (8-day). As shown in Figure 4 (*view EVI for Tchizalamou 2006*) the data from some years had a large variance between consecutive composites and there were cases where the outlier filter proved insufficient, greatly affecting the results [Figure 4, *view EVI for Mongu 2008*]. When removing the shown outliers from the Mongu data set the R^2 increased drastically; from 0.04 to 0.89 ($P < 0.001$).

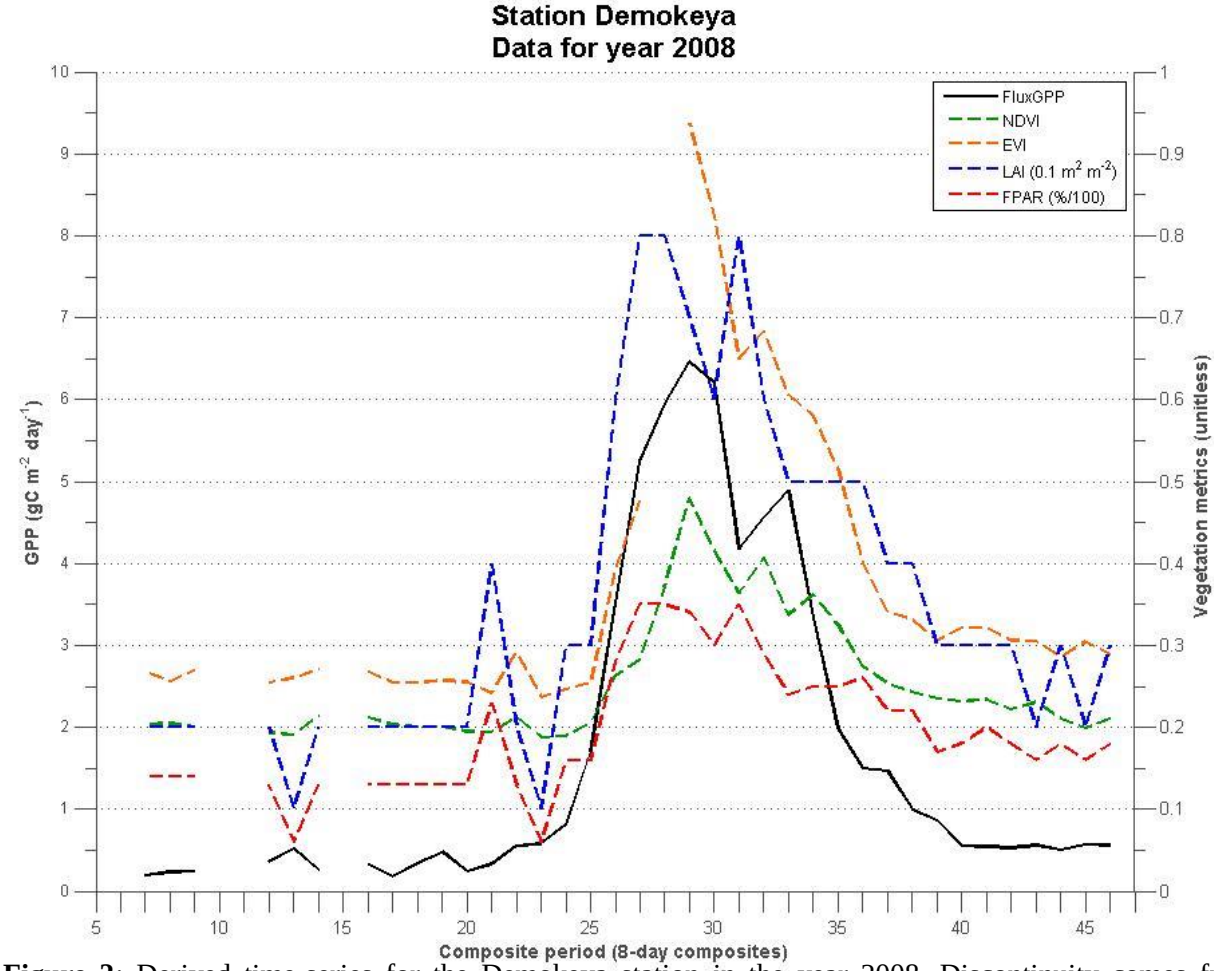

Figure 2*:* Derived time-series for the Demokeya station in the year 2008. Discontinuity comes from missing data. Given are Flux Tower estimated GPP, NDVI (*Normalized Difference Vegetation Index*), EVI (*Enhanced Vegetation Index*), LAI (*Leaf Area Index*) and FPAR (*Fraction of Photosynthetically Active Radiation*).

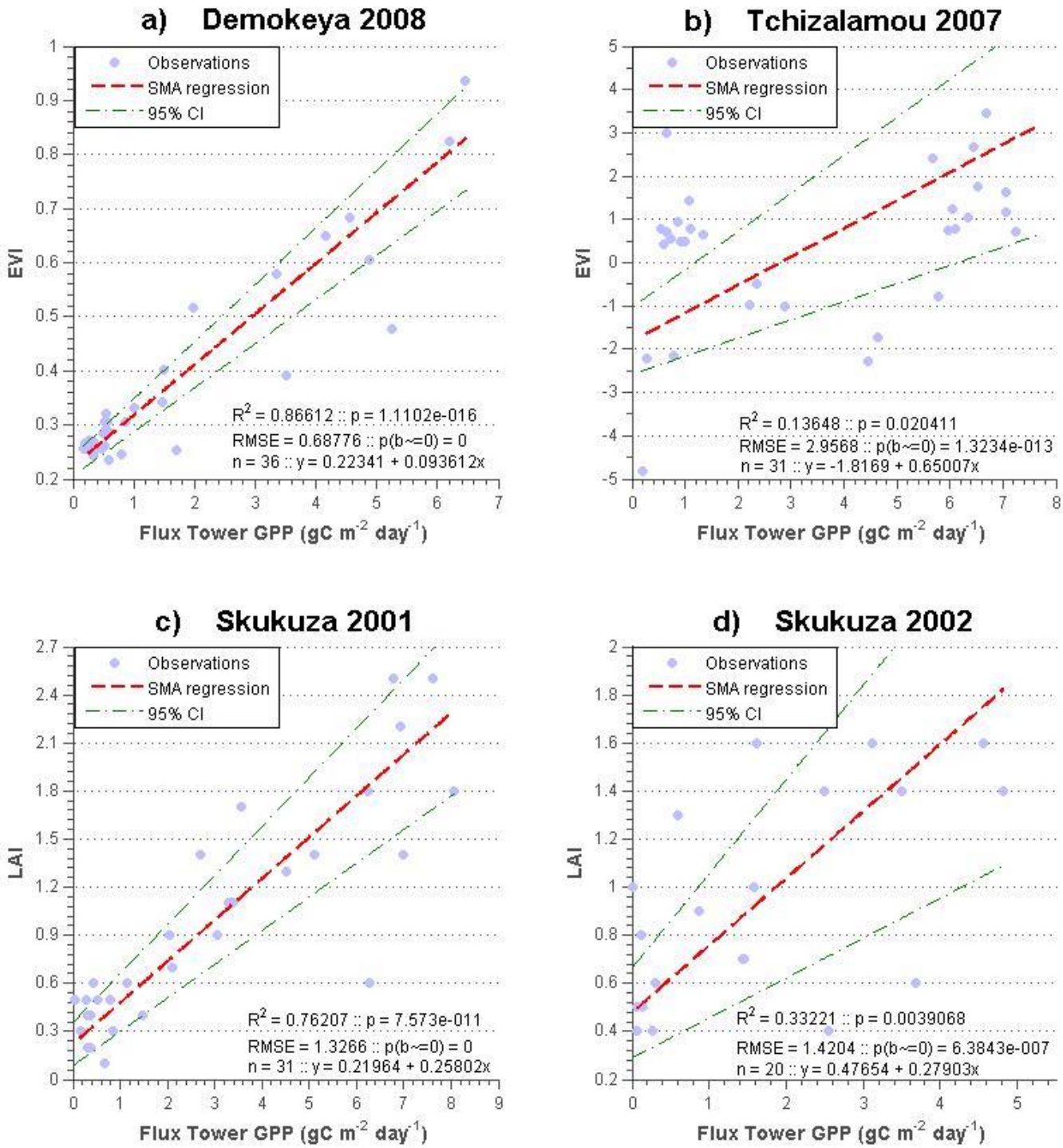

Figure 3*:* EVI (*Enhanced vegetation index*, 8-day) plotted against Flux Tower estimated GPP for the Demokeya 2008 (a) and the Tchizalamou 2007 (b), as well as LAI (*Leaf Area Index*, 8-day) against Flux Tower estimated GPP for the Skukuza year 2001 and 2002. Given are the observations, the Standardized Major Axis Regression line and the 95% confidence interval (CI) for that regression, the coefficient of determination (R^2), the significance thereof (p), RMSE (gC m^{-2} day^{-1}), number of observations (n), the regression model equation (y) and the significance that the slope of the regression is not zero (p(b~=0))

Table 2: Minimum and maximum coefficient of determination (R^2) for each station, vegetation metric and composite size. Given are the site name, min (minimum value), max (maximum value), NDVI (*Normalized Difference Vegetation Index*), EVI (*Enhanced Vegetation Index*), LAI (*Leaf Area Index*) and FPAR (*Fraction of Photosynthetically Active Radiation*) for each composite period. Within parenthesis is given the R^2 when observations outside of the vegetation period were removed by a threshold value (GPP estimate > 0.6 gC m^{-2} day^{-1}), for the respective vegetation metric and composite size. Records with '-' in the 'Max' column and with a value in the 'Min' column indicates that the station only had one year of data in the analysis and hence that 'Max' = 'Min', where the records for both the 'Min' and 'Max' columns are '-' there was no data.

\Stations	**Agoufou**		**Bontioli**		**Demokeya**		**Malopeni**		**Maun**		**Mongu**		**Skukuza**		**Tchizalamou**	
Metrics	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
NDVI	0.83	-	0.55	-	0.62	0.83	0.75	-	-	-	0.13	0.74	0.24	0.74	0.32	0.49
8-day	(-)	(-)	(0.18)	(-)	(0.61)	(0.69)	(-)	(-)	(-)	(-)	(0.13)	(0.72)	(0.12)	(0.80)	(0.28)	(0.28)
EVI	0.76	-	0.76	-	0.53	0.87	0.73	-	-	-	0.04	0.06	0.07	0.73	0.14	0.19
8-day	(-)	(-)	(0.54)	(-)	(0.53)	(0.78)	(-)	(-)	(-)	(-)	(0.03)	(0.06)	(0.01)	(0.76)	(0.07)	(0.07)
LAI	0.77	-	0.54	-	0.67	0.86	0.51	-	0.81	0.81	0.47	0.53	0.28	0.89	0.45	0.53
8-day	(-)	(-)	(0.16)	(-)	(0.53)	(0.86)	(-)	(-)	(-)	(-)	(0.47)	(0.51)	(0.37)	(0.79)	(0.43)	(0.43)
FPAR	0.78	-	0.61	-	0.62	0.78	0.66	-	0.65	0.65	0.34	0.48	0.26	0.76	0.48	0.54
8-day	(-)	(-)	(0.23)	(-)	(0.49)	(0.78)	(-)	(-)	(-)	(-)	(0.34)	(0.43)	(0.28)	(0.68)	(0.46)	(0.46)
NDVI	0.87	-	0.97	-	0.69	0.90	0.87	-	-	-	0.46	0.81	0.50	0.78	0.38	0.85
16-day	(-)	(-)	(-)	(-)	(0.69)	(-)	(-)	(-)	(-)	(-)	(0.46)	(0.77)	(0.55)	(0.62)	(0.32)	(0.32)
EVI	0.79	-	0.80	-	0.69	0.87	0.79	-	-	-	0.28	0.81	0.53	0.96	0.23	0.37
16-day	(-)	(-)	(-)	(-)	(0.69)	(-)	(-)	(-)	(-)	(-)	(0.28)	(0.76)	(0.56)	(0.89)	(0.15)	(0.15)
LAI	0.85	-	0.92	-	0.70	0.92	0.68	-	0.80	0.80	0.65	0.85	0.53	0.90	0.56	0.86
16-day	(-)	(-)	(-)	(-)	(0.92)	(-)	(-)	(-)	(-)	(-)	(0.60)	(0.85)	(0.60)	(0.70)	(0.53)	(0.53)
FPAR	0.83	-	0.92	-	0.65	0.90	0.82	-	0.58	0.58	0.71	0.89	0.44	0.75	0.59	0.83
16-day	(-)	(-)	(-)	(-)	(0.90)	(-)	(-)	(-)	(-)	(-)	(0.63)	(0.89)	(0.55)	(0.63)	(0.56)	(0.56)
NDVI	0.77	-	0.96	-	0.89	0.97	0.94	-	-	-	0.50	0.93	0.59	0.92	0.41	0.86
32-day	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0.50)	(0.93)	(0.64)	(0.65)	(0.56)	(0.56)
EVI	0.85	-	0.90	-	0.82	0.85	0.83	-	-	-	0.35	0.89	0.67	0.97	0.16	0.16
32-day	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0.35)	(0.88)	(0.62)	(0.67)	(0.07)	(0.07)
LAI	0.81	-	0.94	-	0.81	0.90	0.75	-	0.89	0.89	0.73	0.94	0.73	0.88	0.76	0.84
32-day	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0.73)	(0.94)	(0.78)	(0.82)	(0.82)	(0.82)
FPAR	0.77	-	0.98	-	0.79	0.84	0.91	-	0.72	0.72	0.81	0.97	0.60	0.83	0.74	0.81
32-day	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0.81)	(0.97)	(0.71)	(0.73)	(0.81)	(0.81)

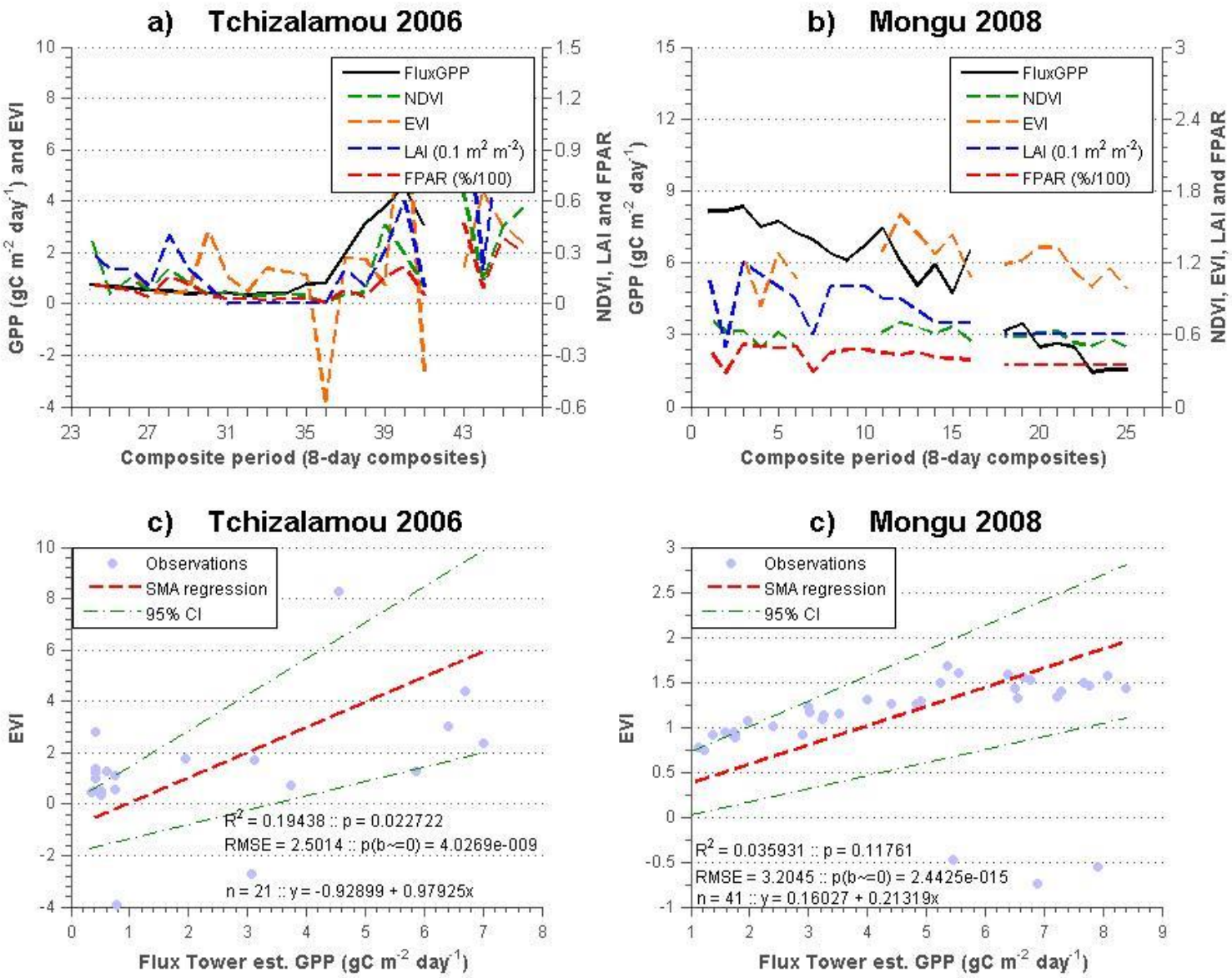

Figure 4: GPP (*Gross Primary Production*), NDVI (*Normalized Difference Vegetation Index)*, EVI (*Enhanced Vegetation Index*), LAI (*Leaf Area Index*), and FPAR (*Fraction of Photosynthetically Active Radiation*) for Tchizalamou 2006 (a) and for the Mongu 2008 (b). EVI versus GPP for the same sites/years (c, d) where the observations, the Standardized Major Axis Regression line and the 95% confidence interval (CI) for that regression, the coefficient of determination (R^2), the significance thereof (p), RMSE (gC m^{-2} day^{-1}), number of observations (n), the regression model equation (y) and the significance that the slope of the regression model is not zero (p(b~=0)), are given.

4.1 Average correlation between MODIS products and Flux Tower GPP per year and station

The results of this part of the analysis attributes LAI as the MODIS product that best estimate Flux Tower GPP intra-annually, explaining 63% of the variance [Table 3], this followed by NDVI at 51% and FPAR at 52% and EVI explaining 49% of the variance. The correlations get markedly higher with increased composite size, e.g. the coefficient of determination for NDVI goes from 0.51 to 0.71 when aggregating to 16-day composite and from 0.71 to 0.76 when aggregating to 32-day composites. Compositing also leads to a lower number of observations and a larger confidence interval.

Removing the Agoufou and Mongu stations not classified as savanna [Table 1] improved the average correlation for half of the vegetation metrics [Table 3], i.e. 8-day R^2 for EVI went from 0.49 to 0.53 and 8-day R^2 for LAI went from 0.63 to 0.64, whereas FPAR was unchanged. The ability of the vegetation metrics to explain the variance of the Flux Tower measured GPP decreased on an average approximately 8% when the observations judged to fall temporally outside of the vegetation period were removed, i.e. when only observations within the vegetation period were used; e.g. the coefficient of determination (R^2) for EVI using all stations and only vegetation period observations was 0.41 compared to 0.49 when using all observations.

The variance in performance was also notable here, e.g. with an average R^2 of 0.17 for EVI at the Tchizalamou station while for the Demokeya station an average R^2 of 0.74 was had for EVI, when taking the average for each of the stations among the years.

Table 3: Average coefficient of determination (R^2) of all years and sites for each MODIS product and composite period; values in parenthesis illustrate average Root Mean Square Error (RMSE, gC m^{-2} day^{-1}). Given are number of stations (n), NDVI (*Normalized Difference Vegetation Index*), EVI (*Enhanced Vegetation Index*), LAI (*Leaf Area Index*) and FPAR (*Fraction of Photosynthetically Active Radiation)* for each composite period. For the columns with 'veg. period' a primitive filter was applied aimed at removing values unlikely to reside within the growth season (Flux Tower Gross Primary Production estimate above 0.6 gC m^{-2} day^{-1}). The asterisks indicate p value < 0.05 (*) and p value < 0.001(**).

	All (n = 8)	**All ('veg. period')**	**Savanna (n = 6)**	**Savanna ('veg. period')**
NDVI 8-day	0.51*(1.72)	0.45*(1.89)	0.50*(1.72)	0.45*(1.85)
EVI 8-day	0.49*(1.76)	0.41 (1.97)	0.53*(1.62)	0.50*(1.74)
LAI 8-day	0.63**(1.40)	0.56*(1.60)	0.64**(1.36)	0.58*(1.55)
FPAR 8-day	0.52*(1.66)	0.50*(1.72)	0.52*(1.64)	0.53*(1.64)
NDVI 16-day	0.71*(1.08)	0.57*(1.49)	0.71*(1.04)	0.55*(1.48)
EVI 16-day	0.67*(1.24)	0.57*(1.48)	0.69*(1.17)	0.59*(1.39)
LAI 16-day	0.79*(0.92)	0.70*(1.17)	0.79*(0.88)	0.69*(1.16)
FPAR 16-day	0.71*(1.06)	0.69*(1.19)	0.68*(1.07)	0.66*(1.23)
NDVI 32-day	0.76*(0.92)	0.66*(1.22)	0.76*(0.91)	0.62*(1.26)
EVI 32-day	0.70*(1.16)	0.52 (1.56)	0.70*(1.15)	0.45 (1.67)
LAI 32-day	0.84*(0.78)	0.82*(0.87)	0.85*(0.74)	0.81*(0.84)
FPAR 32-day	0.79*(0.87)	0.80*(0.87)	0.77*(0.88)	0.75*(0.96)

4.2 Average correlation between MODIS products and Flux Tower GPP per station.

The overall results when pooling all observations for each station regardless of time was decidedly better [Table 4] than for the analysis where all years got equal influence. The R^2 for NDVI increased from 0.51 to 0.57, from 0.49 to 0.61 for EVI, and from 0.52 to 0.59 for FPAR, whereas LAI decreased from 0.63 to 0.62. However there was only a slight overall improvement for the 16-day composites, from an average R^2 of 0.72 to 0.74, and no improvement for the 32-day composite when using the data from all stations.

For the vegetation period correlation an overall decrease was observed between the 4.1 analysis and this part of the analysis, with R^2 for NDVI dropping from 0.45 to 0.36, LAI from 0.56 to 0.44 and FPAR from 0.50 to 0.40, for EVI however, there was an increase from 0.41 to 0.45.

Removing the two non-savanna stations (Agoufou and Mongu) resulted in a slight decrease of R^2, i.e. NDVI from 0.57 to 0.51, EVI from 0.61 to 0.58, LAI from 0.62 to 0.58 and FPAR from 0.59 to 0.57. Also notable was the large variance in between sites when pooling the observations for each site, e.g. the R^2 for EVI for the Tchizalamou site was 0.11 whereas the R^2 for the pooled observations at the Bontioli site was 0.86 for EVI. Notably Tchizalamou and Skukuza explained on average about 35-40% of the variance between the vegetation metrics and Flux Tower derived GPP, whereas stations like Demokeya and Mongu explained on an average 50-70% of the variance.

Table 4: Average coefficient of determination (R^2) of all sites for each MODIS product and composite period; values in parenthesis illustrate average Root Mean Square Error (RMSE, gC m^{-2} day^{-1}). Given are number of stations (n), NDVI (*Normalized Difference Vegetation Index*), EVI (*Enhanced Vegetation Index*), LAI (*Leaf Area Index*) and FPAR (*Fraction of Photosynthetically Active Radiation)* for each composite period. For the columns with 'veg. period' a primitive filter was applied aimed at removing values unlikely to reside within the growth season (Flux Tower Gross Primary Production estimate above 0.6 gC m^{-2} day^{-1}). The asterisks indicate p value < 0.05 (*) and p value < 0.001(**).

	All (n = 8)	**All (veg. period')**	**Savanna (n = 6)**	**Savanna (veg. period')**
NDVI 8-day	0.57**(1.73)	0.36*(2.18)	0.51*(1.93)	0.33*(2.27)
EVI 8-day	0.61*(1.54)	0.45*(1.93)	0.58*(1.62)	0.44 (1.95)
LAI 8-day	0.62**(2.62)	0.44 (2.96)	0.58**(1.88)	0.43 (2.43)
FPAR 8-day	0.59**(3.93)	0.40*(4.61)	0.57*(2.84)	0.40*(3.64)
NDVI 16-day	0.76**(1.13)	0.57*(1.48)	0.75*(1.12)	0.55* (1.47)
EVI 16-day	0.65**(1.51)	0.49*(1.68)	0.61*(1.61)	0.45 (1.73)
LAI 16-day	0.79*(1.06)	0.64**(1.32)	0.79*(1.04)	0.63*(1.31)
FPAR 16-day	0.75**(4.04)	0.58**(4.86)	0.72*(3.27)	0.54*(3.37)
NDVI 32-day	0.78**(1.05)	0.66*(1.20)	0.79*(1.01)	0.63*(1.20)
EVI 32-day	0.67 (1.37)	0.46 (1.67)	0.61 (1.52)	0.37 (1.80)
LAI 32-day	0.84*(0.88)	0.79**(0.96)	0.85*(0.82)	0.77*(0.92)
FPAR 32-day	0.80*(4.24)	0.78**(4.95)	0.79 (3.32)	0.65*(3.26)

4.3 Correlation using all observations for all the savanna sites.

Expanding on the notion that sites classified as savanna sites are comparable the correlation using all observations for all savanna sites was calculated, as can be seen in Figure 5 this resulted in a low coefficient of determination of 0.21 for NDVI, 0.18 for EVI, 0.31 for LAI and 0.24 for FPAR.

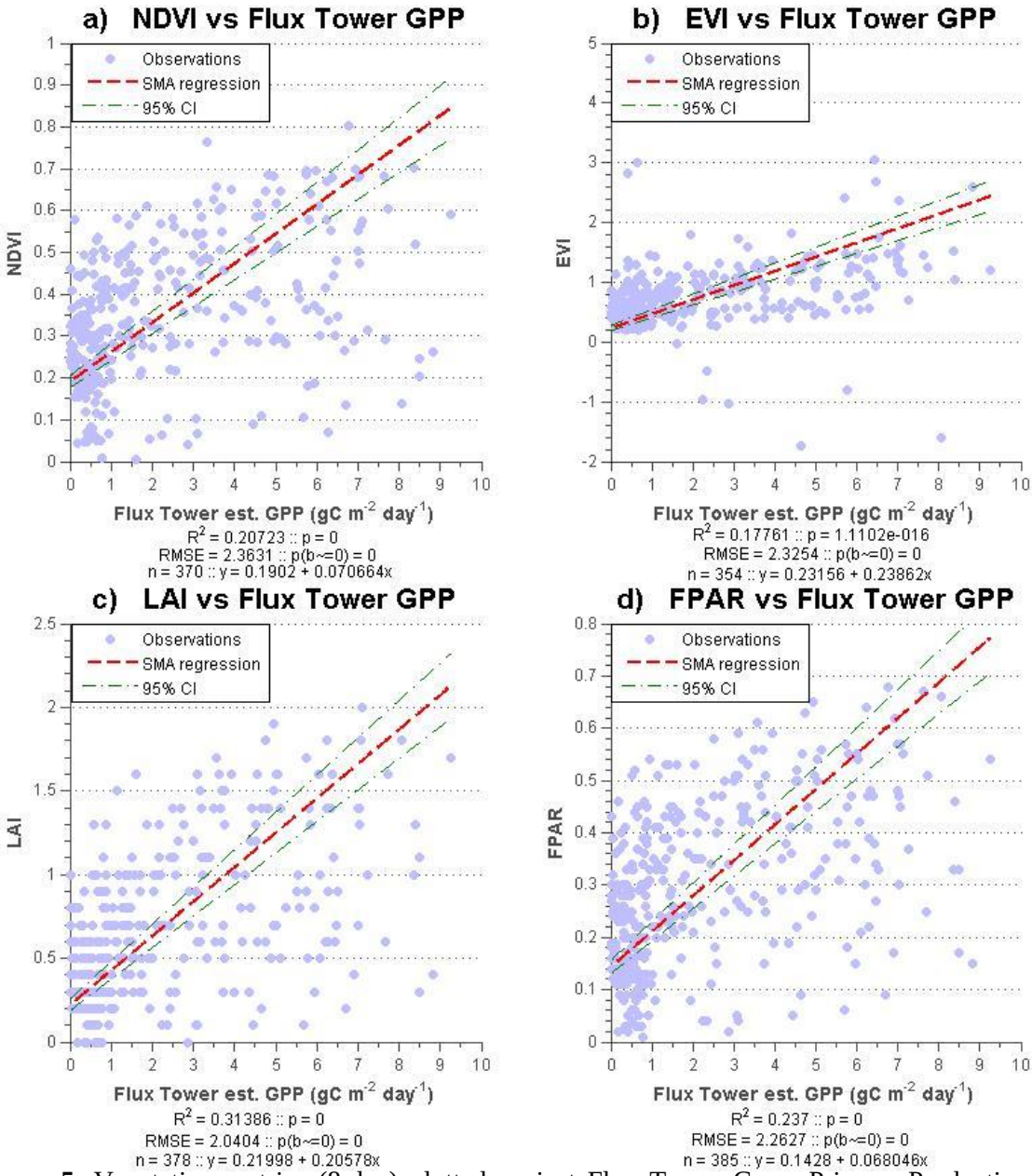

Figure 5: Vegetation metrics (8-day) plotted against Flux Tower Gross Primary Production using all observations available for the ecosystem type savanna. This for NDVI (*Normalized Difference Vegetation Index*, a), EVI (*Enhanced Vegetation Index*, b), LAI (*Leaf Area Index*, c) and FPAR (*Fraction of Photosynthetically Active Radiation*, d). Given are also the observations, the Standardized Major Axis Regression line and the 95% confidence interval (CI) for that regression, the coefficient of determination (R^2), the significance thereof (p), RMSE (gC m^{-2} day^{-1}), number of observations (n), the regression model equation (y) and the significance that the slope of the regression model is not zero (p(b~=0)).

5. DISCUSSION AND LIMITATIONS

5.1 Variation in performance – indicating site dependence?

The variance in performance of the different vegetation metrics observed in the results [Table 2] may indicate a site dependency of performance, e.g. with EVI on an average explaining 17% of the Flux Tower estimated GPP variance at the Tchizalamou station while for the Demokeya station explaining 74% of the Flux Tower estimated variance. This observation is supported by Sjöström *et al.* (2011) whose results also displayed a large variance in performance in between sites e.g. $R^2 = 0.49$ for Skukuza and $R^2 = 0.90$ for Tchizalamou using $EVI \times PAR$; the differences in ability to explain Flux Tower estimated GPP for EVI at the site Tchizalamou between this study and that of Sjöström *et al.* (2011) will be touched upon further down in this discussion. Hashimoto et al. (2012) also viewed site dependence in the performance of EVI over their forested sites, lending further merit to the observed variance in performance and that EVI may be site dependent. As indicated our results show a comparable variance in performance for the other vegetation metrics as well, e.g. 8-day NDVI R^2 for the Tchizalamou site varying between 0.32-0.49 compared to an R^2 varying between 0.62-0.83 for 8-day NDVI at the Demokeya site, extending the notion of site dependency beyond the EVI metric. Variation in performance can be viewed between years of data when observing a single station as well, which coupled with a low sample of years makes conclusions of site dependency precarious.

5.2 LAI as best estimator - of intra-annual GPP variation

The results of the 4.1 analysis identified LAI as the best intra-annual estimate of GPP for which no other studies lending support to has been found; however, LAI has been known to have a strong correlation to GPP for LAI < 4 (Chapin *et al.* 2011), which is true for all the sites in this study but not necessarily for all sites in Hashimoto *et al.* (*2012*). All the while NDVI and EVI are known to be sensitive to the characteristics of arid areas (Sjöström *et al.* 2011). Also Privette *et al.* (2002) concluded that the MODIS LAI product does well for woodland/savanna, so the possibility of LAI surpassing the other vegetation metrics stemming from a combination of these factors arose. However Sjöström *et al.* (2011) and Hashimoto *et al.* (2012) observed EVI as a good predictor of intra-annual variation, further discussed in the next paragraph. Hashimoto *et al.* (2012) also claimed that LAI possesses an inability to respond to short term stresses and notes that as a reason for a poor performance in the estimation of the intra-annual variation of GPP. However since few studies comparing the performance of LAI in predicting variations in Flux Tower GPP has been made no conclusions can be made in this case; it is not unlikely that it performs as viewed here.

5.3 Performance of EVI – compared to other studies

The results of the 4.1 analysis indicated a poor overall performance of EVI whereas Hashimoto *et al.* (2012) brought forth EVI as the best estimator of intra-annual variation of GPP for their forest sites and Sjöström *et al.* (2011) observed a good correlation between EVI and Flux Tower GPP for a series of semi-arid sites. The results of Sjöström *et al.* (2011) was of particular interest to this study due to the similarities in study area, and when closer viewing the methodology of Sjöström *et al.* (2011) two things stood out in comparison to the methodology of this study, which was that they use two years of data for each site i.e. each site had similar temporal scale representation and that they pre-process the data using the TIMESAT tool (Jönsson and Eklundh 2002; 2004). As can be viewed in Sjöström *et al.* (2011, *Fig. 2.*) the TIMESAT tool smooth the MODIS data markedly which is especially pronounced for the Mongu and Tchizalamou sites, sites that in this study displayed fluctuating MODIS data [Figure 3] and gave a poor result, e.g. 8-day EVI at Mongu varying in between 0.04-0.06 and in between 0.14-0.19 for the

Tchizalamou site, and indeed, with a lower tolerance for outliers we saw a drastically improved R^2 for Mongu. Notably the coefficient of determination derived from the TIMESAT smoothed EVI and Flux Tower estimated GPP for the station Tchizalamou by Sjöström *et al.* (2011) was markedly higher than the results from that station in this study. This may merit the use of tools like TIMESAT to improve performance and minimize the influence of error laden observations.

It was noted that when calculating the R^2 for each year the amount of observations and the part of the year those observations covered played an important role in each R^2 value received, which was taken into account in this study. But temporal representation of each site was not accounted for; which proves to be a mistake since when using the methodology of section 4.1 each year, regardless of what station it appertains to, gets equal influence. This becomes problematic since this study has a temporal representation variance that is quite large, e.g. Agoufou with one year of data versus Skukuza with nine years of data; this is especially problematic under the premise that these vegetation metrics performs decidedly different between sites. Notably for the station Skukuza, which was represented with more years of data, a poor performance across the vegetation metrics was observed, e.g. an average R^2 for FPAR of 0.45 for all years, which also can be observed for EVI in Sjöström *et al.* (2011). The importance of comparable temporal representation is further supported by the fact that the correlation improved markedly when each site got equal influence (section 4.2). But the highest influence on the overall performance of EVI (8-day) in section 4.1 seems to stem from the bad performance for some stations and years, e.g. Tchizalamou and Mongu shown in Table 2.

5.4 Ability to handle seasonality

Most of the sites displayed a strong seasonality, clearly viewed in the data [Figure 1] which created clusters of data points [Figure 3, *Demokeya 2008*], and the question arose how well the vegetation metrics handled this seasonality. This is where the concept of extracting the growth season (*as defined by a threshold value of Flux Tower GPP*) came in. It was reasoned that these clusters of non-growing season could have a profound impact on the correlation between the data sets, while pertaining a larger sensitivity to noise. The results show that isolating the vegetation period lends a weaker correlation (*explaining on an average ~8% less of the variance*) and due to the decreased sample size the probability of an accurate assessment is lower. Notable is that the coefficient of determination (R^2) for EVI using all stations and only vegetation period observations was 0.41 compared to 0.49 using all observations. This is also decidedly lower than the R^2 of 0.57 that Sjöström *et al.* (2011) got as a result when they used similar constraints. However the stations used differ in that this study used the stations Agoufou, Bontioli, Demokeya, Malopeni, Maun, Mongu, Skukuza and Tchizalamou whereas Sjöström *et al.* (2011) used the stations Demokeya, Maun, Mongu, Skukuza, Tchizalamou, Wankama Fallow and Wankama Millet; and this at a different temporal scale. The contrast of the increased performance when pooling the observations in the 4.2 analysis compared to the 4.1 analysis towards that of the decrease when viewing the vegetation period observations only, may indicate that non-growing season observations leads to an overestimation of the performance of the vegetation metrics.

While it was believed that these values measured during the non-vegetation season could portray little vegetation and show higher susceptibility to noise, or rather face a larger relative impact for these low measurement values. It was ruled that removing these data points entirely, even though the risk of these observation portraying anything but vegetation, would remove an interesting chance for analysis and furthermore could not be motivated without deeper in situ knowledge and of MODIS quality. Therefore naught more is derived from this part of the analysis than the notion that the vegetation metrics may be less proficient in mapping these often distinct peaks of vegetation and that the dry season may masks the true ability of the vegetation metrics. It is also problematic to use the coefficient of determination to view the relations between sets like these where you have a skewed distribution.

5.5 Potential improvements and sources of error

The correlations get markedly higher with increased composite size, e.g. the coefficient of determination for NDVI goes from 0.51 to 0.71 [Table 3] when aggregating to 16-day composite and from 0.71 to 0.76 when aggregating to 32-day composites [Table 3]. This aggregation has in a sense a similar function as that of the TIMESAT tool as it results in a curve that is connected to the original observations, or rather the trend thereof, however since this aggregation for the MODIS data uses the MVC method any large positive deviation will remain and there are also issues with where the missing data occurs. Also when compositing like this the number of observations decreases and the confidence interval gets larger. Hence using the MVC method could be a potential source of error and reduces the likelihood that the resulting R^2 represents the true R^2 value. An improvement to this study could therefore be to use a tool like that of TIMESAT which Sjöström *et al.* (2011) showed improves the accuracy of intra-annual estimation.

As seen in Figure 4 (*EVI for Mongu 2008*) there are a couple of values not taken by the outlier filter that deviates from the rest of the observations, this observation was traced back to the MOD09A1 source where the quality value for that pixel indicated bad quality for band 3. This demonstrates the importance of utilizing the quality indicators to avoid some potential issues and assure a higher quality outcome, but also that a better outlier control would likely improve the results. Therefore this study could be improved by using the available quality information, or at the least give an indication of potential errors in measurements. When replacing cloud flagged composites perhaps a window using the average of the two adjacent values would yield more accurate results than that of a MVC window.

Furthermore weights could be used to circumvent the temporal issue discussed in the segment about the performance of EVI, vegetation type and site characteristics received too little attention in this study and a prevailing issue is that of sample size, more data would be good.

5.6 Concluding thoughts

The analysis in section *4.2* was aimed at simulating better coverage, but interpretation should be done with caution because it may lead to overrepresentation of some features depending on the look of each set. That the correlation improved could however imply that we are underestimating the ability of the metrics to estimate fluctuations in GPP, either because of small sample sizes or via temporal misrepresentation as in the discussion about the section *4.1* above; or it could simply mean that the phenomenon discussed of the non-growing season observations get to exert an even greater influence, however as it seems that would rather lead to an overestimation.

The section *4.3* was performed knowing interpretation may be convoluted due to the known diversity of the sites, however, the results (R^2 *~0.18-0.31*) could imply, as previously touched upon, that these metrics are better at measuring some characteristics than others; Hence the difference between sites. We are after all plotting each satellite observation paired with each station measurement regardless of location and time, which given perfect unison between measurement tool and reality, should not matter.

The results from this study compared to e.g. that of Sjöström *et al.* (2011) showed that including meteorological factors controlling plant carbon assimilation and respiration improve the ability of models to estimate GPP. These meteorological factors may vary in importance between sites, e.g. water availability is highly important for carbon assimilation in arid areas (Hickler *et al.* 2005), which lends the discussion of whether a changed approach towards algorithms estimating GPP applying different models depending on site characteristics. Arguably this calls for extensive evaluation of different models performance for different biomes, for higher resolution of data and more narrowly classified biomes; all foreseeably possible in the scientific community of today.

6. CONCLUSIONS

This study concludes that the vegetation metrics compared perform modestly when observed as a whole; however they display a large variance which means that they perform quite well for some sites and this variance may also indicate site-dependency for the vegetation metrics. An improvement was observed when classifying after savanna, but due to the low difference in number of participatory years of data (3 years) in between using only savanna sites and using all sites and years conclusions are precarious.

Since the results points towards these vegetation metrics being site dependent and their performance modest, simplistic regression models utilizing these individual metrics alone are considered to be unwise to scale beyond site local. While little can be concluded by the discussion on the topic about the ability to handle the seasonality in this study it is apparent that the non-vegetation period observations had a great influence on the derived coefficient of determination.

However the relationships are undeniable and the potential shown in other studies using ancillary information indicates that remotely sensed data holds great potential of in the future providing large scale estimations of GPP. There is however much to be done in the accuracy of the measurement tools, creation of models and validation thereof. It is my belief that the use of biome specific models combined with a higher resolution of data and a higher specialization classification system bears exciting potentials in this field.

REFERENCES

Archibald, S.A., Kirton, A., van der Merwe, M.R., Scholes, R.J., Williams, C.A., Hanan, N., 2009. Drivers of inter-annual variability in Net Ecosystem Exchange in a semi-arid savanna ecosystem, South Africa. *Biogeosciences, 6*, 251-266.

Asrar, G., Fuchs, M., Kanemasu, E.T., Hatfield, J.L., 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. *Agronomy Journal, 76*, 300-306.

Baldocchi, D. D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future. *Global Change Biology*, 9, 479−492, doi:10.1046/j.1365-2486.2003.00629.x.

Baldocchi, D., Falge, E., Gu, L.H., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X.H., Malhi, Y., Meyers, T., Munger, W., Oechel, W., U, K.T.P., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society, 82*, 2415-2434.

Bombelli, A., Henry, M., Castaldi, S., Adu-Bredu, S., Arneth, A., de Grandcourt, A., et al., 2009. An outlook on the Sub-Saharan Africa carbon balance. *Biogeosciences*, 6, 2193−2205, doi:10.5194/bg-6-2193-2009.

Fensholt, R., Sandholt, I., Rasmussen, M.S., 2004. Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sensing of Environment, 91*, 490-507.

Hashimoto, H., Wang, W., Milesi, C., White, M. A., Ganguly, S., Gamo, M., Hirata, R., Myneni, R. B., Nemani, R. R., 2012. Exploring Simple Algorithms for Estimating Gross Primary Production in Forested Areas from Satellite Data. *Remote Sensing*, 4, 303-326, doi: 10.3390/rs4010303

Hickler, T., Eklundh, L., Seaquist, J.W., Smith, B., Ardö, J., Olsson, L., Sykes, M.T., Sjöström, M., 2005. Precipitation controls Sahel greening trend. *Geophysical Research Letters*, 32, doi: 10.1029/2005GL024370.

Huete, A.R., 1988. A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*, 295-309.

Huete, A.R., Liu, H.Q., Batchily, K., van Leeuwen, W., 1997. A comparison of vegetation indices global set of TM images for EOS-MODIS. *Remote Sensing of Environment, 59*, 440-451.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, 83*, 195-213.

Jönsson, P. and Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data, IEEE T. *Geosci. Remote, vol* 40, 1824–1832, doi:10.1109/TGRS.2002.802519.

Jönsson, P. and Eklundh, L., 2004. TIMESAT – a program for analyzing time-series of satellite sensor data, *Comput. Geosci.*, 833–845, doi:10.1016/j.cageo.2004.05.006.

Knyazikhin, Y., Martonchik, J. V., Myneni, R. B., Diner, D. J., Running, S. W., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *J. Geophys. Res.*, 103, 32257-32275.

Lee, X., Massman, W. and Law, B.E., 2004. Handbook of micrometerology. A guide for surface flux measurement and analysis. *Kluwer Academic Press, Dordrecht*, 250 pp.

Massman, W.J and Lee, X., 2002. Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges. *Agricultural and Forest Meteorology*, 113(1-4):121-144.

Myneni, R. B., Nemani, R. R. and Running, S. W., 1997. Estimation of Global Leaf Area Index and Absorbed Par Using Radiative Transfer Models. IEEE T. *Geoscience and Remote sensing*, *vol* 35, *issue* 6, 1380-1393,.

Papale, D. and Valentini, R., 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology*, 9, 525−535, doi:10.1046/j.1365-2486.2003.00609.x.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., Yakir, D., *2006.* Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation. *Biogeosciences*, *3*, 571-583.

Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., 2007. Climate change 2007: impacts, adaptation and vulnerability. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. *Cambridge: Cambridge University Press*.

Perry, C. R. and L. F. Lautenschlager., 1984. Functional Equvalence of Spectral Vegetation Indices. *Remote Sensing of Environment*, Vol. 14, pp. 169-182.

Privette, J. L., Myneni, R. B., Knyazikhin, Y., Mukelabai, M., Roberts, G., Tian, Y., Wang, Y., Leblanc, S.G., 2002. Early spatial and temporal validation of MODIS LAI product in the Southern Africa Kalahari. *Remote Sensing of Environment*, 83, 232−243.

Reed, B. C., J. Brown, D. VanderZee, T. Loveland, J. Merchant, and D. Ohlen., 1994. Measuring Phenological Variability from Satellite Imagery. *Journal of Vegetation Science*, Vol. 3, pp. 703-714.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránkov, K., Ilvesnemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A. Loustau, D. Matteucci, G., Meyers, T., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. *Global Change Biol.*, *11*, 1424-1439.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W., 1973. Monitoring Vegetation Systems in the Great Plains with ERTS. *In Third ERTS Symposium* (NASA Special Publication SP-3511). Pp. 309-317.

Running, S.W., Nemani, R.R., Peterson, D.L., Band, L.E., Potts, D.F., Pierce, L.L., Spanner, M.A, 1989. Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation. *Ecology*, *70*, 1090-1101.

Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Flanagan, L.B., Goldstein, A.H., Hollinger, D.Y., Misson, L., Monson, R.K., Oechel, W.C., Schmid, H.P., Wofsy, S.C., Xu, L.K., 2006. On the use of MODIS EVI to assess gross primary productivity of North American ecosystems. *Journal of Geophysical Research. G. Biogeosciences, 111, G04015.*

Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Bolstad, P. V., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schmid, H. P., Wolfsy, S. C., Xu, L., 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. Remote Sensing of Environment, 112, 1633-1646, doi: 10.1016/j.rse.2007.08.004.

Sjöström, M., Ardö, J., Eklundh, L., El-Tahir, B.A., El-Khidir, H.A.M., Hellström, M., Pilesjö, P., Seaquist, J., 2009. Evaluation of satellite based indices for gross primary production estimates in a sparse savanna in the Sudan. *Biogeosciences, 6*, 129-138, doi:10.5194/bg-6-129-2009.

Sjöström, M., Ardö, J., Arneth, A., Boulain, N., Cappelaere, B., Eklundh, L., de Grandcourt, A., Kutsch, W.L., Merbold, L., Nouvellon, Y., Scholes, R.J., Schubert, P., Seaquist, J., Veenendaal, E.M., 2011. Exploring the potential of MODIS EVI for modeling gross primary production across African ecosystems. *Remote Sensing of Environment*, 115, 1081-1089, doi:10.1016/j.rse.2010.12.013.

Sjöström, M., Zhou, M., Archibald, S., Arneth, A., Cappelaere, B., Falk, U., Grandcourt A., Hanan, N., Kergoat, L., Kutsch, W., Merbold, L., Mougin, E., Nickless, A., Nouvellon, Y., Scholes, B., Veenendaal, E.M., Ardö, J., 2013. Evaluation of MODIS gross primary productivity for Africa using eddy covariance data. *Remote Sensing of Environment, 131, 275-286, doi: 10.1016/j.rse.2012.12.023.*

Schmid, H. P., 2002. Footprint modeling for vegetation atmosphere exchange studies: A review and perspective. *Agricultural and Forest Meteorology*, 113, 159−183, doi:10.1016/S0168-1923(02)00107-7.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L., 2007. Climate change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge: Cambridge University Press*

Tan, B., Woodcock, C. E., Hu, J., Zhang, P., Ozdogan, M., Huang, D., Yang, W., Knyazikhin, Y., Myneni, R.B., 2006. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. *Remote Sensing of Environment*, 105, 98−114, doi:10.1016/j.rse.2006.06.008.

Tian, YH, Zhang, Y, Knyazikhin, Y, Myneni, RB, Glassy, JM, Dedieu, G, Running, SW, 2000. Prototyping of MODIS LAI and FPAR algorithm with LASUR and LANDSAT data. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 38, 5, p2387-p2401

Tucker, C. J., 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, Vol. 8, pp. 127-150.

Turner, D.P., Ritts, W.D., Cohen, W.B., Gower, S.T., Running, S.W., Zhao, M., Costa, M.H., Kirschbaum, A.A., Ham, J.M., Saleska, S.R., Ahl, D.E., 2006. Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment, 102*, 282-292.

Verma, S.B., 1990. Micrometerological methods for measuring surface fluxes of mass and energy. *Remote Sensing Reviews*, 5: 99-115.

Wang, Y., Tian, Y., Zhang, Y., El-Saleous, N., Knyazikhin, Y., Vermote, E., Myneni, R.B., 2001. Investigation of product accuracy as a function of input and model uncertainities: case study with SeaWiFS and MODIS LAI/FPAR Algorithm. *Remote Sensing of Environment*, 78: 296-311.

Xiao, X. M., Braswell, B., Zhang, Q., Boles, S., Frolking, S., Moore, B., 2003. Sensitivity of vegetation indices to atmospheric aerosols: continental-scale observations in Northern Asia, *Remote Sensing of Environment*, 84, 385–392.

Xiao, X., Hollinger, D., Aber, J., Goltz, M., Davidson, E.A., Zhang, Q., Moore, B., III., 2004. Satellite-based modeling of gross primary production in an evergreen needleleaf forest. *Remote Sensing of Environment,* 89, 519-534.

Hardcopy resources:

Brown, Molly E., (2008). *Famine early warning systems and remote sensing data*. Berlin: Springer

Campbell, James B., (2006). *Introduction to remote sensing*. 4th ed. Abingdon: Taylor & Francis

Chapin, III, F. Stuart., Matson, Pamela A., Vitousek, Peter M., (2011). *Principles of Terrestrial Ecosystem Ecology [Elektronisk resurs]*. New York, NY: Springer New York

Eric F. Lambin, Helmut Geist, (2006). *Land-use and land-cover change: local processes and global impacts*. Berlin: Springer

Sjöström, M., 2012. Satellite remote sensing of primary production in semi-arid Africa. *Doctoral thesis:* Department of Physical Geography and Ecosystem Science, Lund University, Tryckeriet i E-huset.

Sokal, Robert R. & Rohlf, F. James, (1995). *Biometry: the principles and practice of statistics in biological research*. 3rd ed. New York: W. H. Freeman

Internet resources:

ORNL DAAC. Available from 2009. MODIS Subsetted Land Products, Collection 5. *Oak Ridge, TN, USA*. Available online: http://daac.ornl.gov/MODIS/modis.html (Accessed on 19th December 2012). Data from selected field site or created subset.

MOD09. Available from 2011. Vermote, E. F., Kotchenova, S. Y., Ray, J. P., MODIS Surface Reflectance User's Guide. *MODIS Land Surface Reflectance Science Computing Facility*. Available online: http://modis-sr.ltdri.org/products/MOD09_UserGuide_v1_3.pdf (Accessed on the 30th of January 2013).

MOD15a. Available from October 9th 2003. Myneni, R.B., Knyazikhin, Y., Glassy, J., Votava, P., Shabanov, N, User's Guide FPAR, LAI (ESDT: MOD15A2) 8-day Composite NASA MODIS Land Algorithm, Available online: http://cybele.bu.edu/modismisr/products/modis/userguide.pdf (Accessed on 21th of February 2013)

MOD15b. Available from April 30th 1999. Knyazikhin, Y., Glassy, J., Privette, J. L., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J. T., Votava, P., Myneni, R.B., Nemani, R. R., Running, S. W., MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document, Available online: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod15.pdf (Accessed on 27th of January 2013)

Student examensarbete (Seminarieuppsatser). Uppsatserna finns tillgängliga på institutionens geobibliotek, Sölvegatan 12, 223 62 LUND. Serien startade 1985. Hela listan och själva uppsatserna är även tillgängliga på LUP student papers (www.nateko.lu.se/masterthesis) och via Geobiblioteket (www.geobib.lu.se)

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 1985. The complete list and electronic versions are also electronic available at the LUP student papers (www.nateko.lu.se/masterthesis) and through the Geo-library (www.geobib.lu.se)

230 Cléber Domingos Arruda (2011) Developing a Pedestrian Route Network Service (PRNS)
231 Nitin Chaudhary (2011) Evaluation of RCA & RCA GUESS and estimation of vegetation-climate feedbacks over India for present climate
232 Bjarne Munk Lyshede (2012) Diurnal variations in methane flux in a low-arctic fen in Southwest Greenland
233 Zhendong Wu (2012) Dissolved methane dynamics in a subarctic peatland
234 Lars Johansson (2012) Modelling near ground wind speed in urban environments using high-resolution digital surface models and statistical methods
235 Sanna Dufbäck (2012) Lokal dagvattenhantering med grönytefaktorn
236 Arash Amiri (2012)Automatic Geospatial Web Service Composition for Developing a Routing System
237 Emma Li Johansson (2012) The Melting Himalayas: Examples of Water Harvesting Techniques
238 Adelina Osmani (2012) Forests as carbon sinks - A comparison between the boreal forest and the tropical forest
239 Uta Klönne (2012) Drought in the Sahel – global and local driving forces and their impact on vegetation in the 20th and 21st century
240 Max van Meeningen (2012) Metanutsläpp från det smältande Arktis
241 Joakim Lindberg (2012) Analys av tillväxt för enskilda träd efter gallring i ett blandbestånd av gran och tall, Sverige
242 Caroline Jonsson (2012) The relationship between climate change and grazing by herbivores; their impact on the carbon cycle in Arctic environments
243 Carolina Emanuelsson and Elna Rasmusson (2012) The effects of soil erosion on nutrient content in smallholding tea lands in Matara district, Sri Lanka
244 John Bengtsson and Eric Torkelsson (2012) The Potential Impact of Changing Vegetation on Thawing Permafrost: Effects of manipulated vegetation on summer ground temperatures and soil moisture in Abisko, Sweden
245 Linnea Jonsson (2012). Impacts of climate change on Pedunculate oak and Phytophthora activity in north and central Europe
246 Ulrika Belsing (2012) Arktis och Antarktis föränderliga havsistäcken

247 Anna Lindstein (2012) Riskområden för erosion och näringsläckage i Segeåns avrinningsområde
248 Bodil Englund (2012) Klimatanpassningsarbete kring stigande havsnivåer i Kalmar läns kustkommuner
249 Alexandra Dicander (2012) GIS-baserad översvämningskartering i Segeåns avrinningsområde
250 Johannes Jonsson (2012) Defining phenology events with digital repeat photography
251 Joel Lilljebjörn (2012) Flygbildsbaserad skyddszonsinventering vid Segeå
252 Camilla Persson (2012) Beräkning av glaciärers massbalans – En metodanalys med fjärranalys och jämviktslinjehöjd över Storglaciären
253 Rebecka Nilsson (2012) Torkan i Australien 2002-2010 Analys av möjliga orsaker och effekter
254 Ning Zhang (2012) Automated plane detection and extraction from airborne laser scanning data of dense urban areas
255 Bawar Tahir (2012) Comparison of the water balance of two forest stands using the BROOK90 model
256 Shubhangi Lamba (2012) Estimating contemporary methane emissions from tropical wetlands using multiple modelling approaches
257 Mohammed S. Alwesabi (2012) MODIS NDVI satellite data for assessing drought in Somalia during the period 2000-2011
258 Christine Walsh (2012) Aerosol light absorption measurement techniques: A comparison of methods from field data and laboratory experimentation
259 Jole Forsmoo (2012) Desertification in China, causes and preventive actions in modern time
260 Min Wang (2012) Seasonal and inter-annual variability of soil respiration at Skyttorp, a Swedish boreal forest
261 Erica Perming (2012) Nitrogen Footprint vs. Life Cycle Impact Assessment methods – A comparison of the methods in a case study.
262 Sarah Loudin (2012) The response of European forests to the change in summer temperatures: a comparison between normal and warm years, from 1996 to 2006
263 Peng Wang (2012) Web-based public participation GIS application – a case study on flood emergency management
264 Minyi Pan (2012) Uncertainty and Sensitivity Analysis in Soil Strata Model Generation for Ground Settlement Risk Evaluation
265 Mohamed Ahmed (2012) Significance of soil moisture on vegetation greenness in the African Sahel from 1982 to 2008
266 Iurii Shendryk (2013) Integration of LiDAR data and satellite imagery for biomass estimation in conifer-dominated forest
267 Kristian Morin (2013) Mapping moth induced birch forest damage in northern Sweden, with MODIS satellite data
268 Ylva Persson (2013) Refining fuel loads in LPJ-GUESS-SPITFIRE for wet-dry areas - with an emphasis on Kruger National Park in South Africa
269 Md. Ahsan Mozaffar (2013) Biogenic volatile organic compound emissions from Willow trees
270 Lingrui Qi (2013) Urban land expansion model based on SLEUTH, a case

study in Dongguan City, China

271 Hasan Mohammed Hameed (2013) Water harvesting in Erbil Governorate, Kurdistan region, Iraq - Detection of suitable sites by using Geographic Information System and Remote Sensing

272 Fredrik Alström (2013) Effekter av en havsnivåhöjning kring Falsterbohalvön.

273 Lovisa Dahlquist (2013) Miljöeffekter av jordbruksinvesteringar i Etiopien

274 Sebastian Andersson Hylander (2013) Ekosystemtjänster i svenska agroforestrysystem

275 Vlad Pirvulescu (2013) Application of the eddy-covariance method under the canopy at a boreal forest site in central Sweden

276 Malin Broberg (2013) Emissions of biogenic volatile organic compounds in a Salix biofuel plantation – field study in Grästorp (Sweden)

277 Linn Renström (2013) Flygbildsbaserad förändringsstudie inom skyddszoner längs vattendrag

278 Josefin Methi Sundell (2013) Skötseleffekter av miljöersättningen för natur- och kulturmiljöer i odlingslandskapets småbiotoper

279 Kristín Agustsdottír (2013) Fishing from Space: Mackerel fishing in Icelandic waters and correlation with satellite variables

280 Cristián Escobar Avaria (2013) Simulating current regional pattern and composition of Chilean native forests using a dynamic ecosystem model

281 Martin Nilsson (2013) Comparison of MODIS-Algorithms for Estimating Gross Primary Production from Satellite Data in semi-arid Africa

282 Victor Strevens Bolmgren (2013) The Road to Happiness – A Spatial Study of Accessibility and Well-Being in Hambantota, Sri Lanka